

Tumor heterogenitás vizsgálata szolid tumorokban újgenerációs DNS szekvenálás segítségével

Doktori értekezés

Pongor Lőrinc Sándor

Semmelweis Egyetem

Patológiai Tudományok Doktori Iskola



Témavezető: Dr. Györffy Balázs, az MTA doktora, tudományos tanácsadó

Hivatalos bírálók: Dr. Kocsis István, Ph.D., egyetemi docens

Dr. Mészáros Bálint, Ph.D., tudományos munkatárs

Szigorlati bizottság elnöke: Dr. Veres Gábor, az MTA doktora, egyetemi tanár

Szigorlati bizottság tagjai: Dr. Alexa Anita, Ph.D., tudományos munkatárs

Dr. Újházy András, Ph.D., egyetemi adjunktus

Budapest

2018

Tartalomjegyzék

1. Rövidítések jegyzéke.....	5
2. Bevezetés.....	11
2.1. A tumorok kialakulását elősegítő molekuláris változások.....	11
2.2. Tumorok mutagenezise és a mutációs mintázatok.....	14
2.3. Karcinogenezis során kialakuló molekuláris változások.....	16
2.4. Az emlőtumorok fő molekuláris jellemzői.....	18
2.5. A melanómák fő molekuláris jellemzői.....	19
2.6. A petefészek-tumorok molekuláris jellemzői.....	19
2.7. Gyógyszeres kezelés, szelekció és gyógyszer rezisztencia.....	20
2.8. A tumor evolúció modelljei.....	24
2.8.1. A klonális szelekció.....	24
2.8.2. A tumor összejt modell.....	25
2.9. Az intra-tumor heterogenitás és vizsgálata.....	25
2.10. Multi-régió szekvenálást alkalmazó vizsgálatok.....	26
2.11. A DNS szekvenálás.....	28
2.11.1. Sanger szekvenálás.....	28
2.11.2. Újgenerációs szekvenálási módszerek.....	29
2.12. Orvosi kutatásban alkalmazott újgenerációs szekvenálási technikák.....	31
2.13. Az újgenerációs szekvenálási adatok elemzése.....	33
2.13.1. A FastQ formátum.....	33
2.13.2. A FastQ adatok feldolgozásának lépései.....	33
2.13.3. Adatok minőségének ellenőrzése és nyírás.....	34
2.13.4. A leolvasások illesztése.....	35
2.13.5. Mutáció és epigenetikai alapú szekvenálási adatok feldolgozása.....	35
2.13.6. Az RNS szekvenálási adatok alap feldolgozása.....	36

2.13.7. A DNS mutációk azonosítása	37
2.13.8. Kópiaszám-változás események azonosítása és a tumor arány meghatározása	38
2.14. A túlélés-elemzés.....	39
3. Célkitűzések	41
4. Módszerek	42
4.1. <i>In vitro</i> sejtvonalas inváziós kísérletek.....	42
4.1.1. Sejtkultúrák.....	42
4.1.1. Gyűrűs inváziós kísérletek	43
4.1.2. DNS izolálás és minőségellenőrzés	44
4.1.3. Sejtvonal-specifikus mutációk igazolása.....	45
4.1.4. Videomikroszkópia.....	46
4.1.5. Sejtmozgás mennyiségi meghatározása.....	46
4.1.6. Célzott szekvenálás Ion Torrent technikával.....	47
4.1.7. Célzott újgenerációs szekvenálási adatok bioinformatikai feldolgozása	47
4.1.8. A szekvenálási lefedettségek összefüggésének <i>in silico</i> vizsgálata a mutációk frekvenciájával.....	47
4.2. Az ovárium-tumor minták újgenerációs szekvenálási adatainak feldolgozása ...	48
4.2.1. Leolvasások illesztése a referencia-genomra	50
4.2.2. Az illesztések előfeldolgozása.....	51
4.2.3. A szomatikus és az öröklött mutációk azonosítása és annotálása	51
4.2.4. A szomatikus mutáció együttes azonosítása.....	53
4.2.5. A leolvasások szűrése és beolvasása az mcaller szoftverben	53
4.2.6. Mutációk keresés aktív régiókban	54
4.2.7. Kópiaszám változás elemzés	55
4.2.8. A mutációs mintázatok meghatározása	55
4.3. A TCGA adatbázis emlődaganatos és nem-kissejtes tüdőrák mintáinak feldolgozása.....	56

4.3.1. Azonosított mutációk annotálása.....	56
4.3.2. Génkifejeződési adatok gyűjtése és normalizálása.....	57
4.3.3. Statisztikai elemzés	57
4.3.4. A G-2-O (Genotype-to-Outcome) elemző rendszer	57
5. Eredmények.....	60
5.1. A sejtvonalak motilitásának hatása az újgenerációs szekvenálás eredményére ..	60
5.1.1. Sejtvonal paraméterek meghatározása monokultúrákkal	60
5.1.2. Az inváziós kísérletek.....	61
5.1.3. Sejtvonal kalibrációs sorok elemzése	64
5.2. A petefészek tumorok multirégiós szekvenálása.....	67
5.2.1. Homológ rekombinációban fontos szerepű gének mutációi.....	67
5.2.2. Kópiaszám-változás elemzése és a tumor-arány számítása.....	68
5.2.3. Szomatikus mutációk azonosítása	70
5.2.4. Egyes tumor régiókban azonosított mutációk számának összehasonlítása ..	72
5.2.5. Szomatikus mutáció mintázatok.....	74
5.3. Túlélés elemzés mutációhoz kapcsolt génkifejeződés változással emlőtumoros betegekben.....	75
5.3.1. Túlélés elemzés az emlőtumor szubtypusokban.....	80
5.3.2. A TP53 génmutációhoz kötött génkifejeződés változás.....	82
5.3.3. Gén ontológia elemzés a TP53 gén mutációjával asszociált génekkel emlő és NSCLC betegekben	83
5.3.4. Túlélés-elemzés sejtciklus szabályozással összefüggő génekkel	85
6. Megbeszélés	88
6.1. A sejtmozgás <i>in vitro</i> követése újgenerációs szekvenálással	88
6.2. Az azonosított mutációk összefüggése a tumor minta méretével petefészek-tumoros betegekben	89
6.3. A Genotype-2-Outcome (G-2-O) elemzőrendszer	90
7. Következtetések.....	92

8. Összefoglalás	94
9. Summary.....	95
10. Irodalomjegyzék	96
11. Saját publikációk jegyzéke	124
11.1. Disszertációhoz kapcsolódó publikációk jegyzéke	124
11.2. Disszertációtól független publikációk jegyzéke	124
12. Köszönetnyilvánítás	126

1. Rövidítések jegyzéke

Rövidítés	Angol kifejezés	Magyar kifejezés
A1CF	Apobec1 Complementation Factor	Apobec1 komplementációs faktor
ABL2	Abl Proto-Oncogene 2	Abl proto-onkogén 2
AFF2	Af4/Fmr2 Family Member 2	Af4/Fmr2 család 2-es tagja
AKT	Rac-Alpha Serine/Treonine-Protein Kinase	Rac-Alfa szerin/treonin fehérje kináz
AMER1	Apc Membrane Recruitment Protein 1	Apc membrán toborzó fehérje 1
APC	Adenomatous Polyposis Coli Protein	Adenomatous Polyposis Coli fehérje
APOBEC	Apolipoprotein B Mrna Editing Enzyme	mRNS módosító enzim B apolipofehérje
ARID1A	At-Rich Interaction Domain 1A	AT-gazdag interakciós 1A Domén
ATM	Ataxia Telangiectasia Mutated	Ataxia Telangiectasia mutált
ATP1A1	Atpase Na ⁺ /K ⁺ Transporting Subunit Alpha 1	ATP-függő Na ⁺ /K ⁺ csatorna Alfa 1 alegység
ATR	Ataxia Telangiectasia And Rad3-Related Protein	Ataxia Telangiectasia és Rad3 kapcsolt fehérje
AUC	Area under the curve	görbe alatti terület
BAM	Binary alignment and mapping	Bináris illeszkedési formátum
BCL11B	B Cell Cll/Lymphoma 11B	B-Sejt Cll/Limfóma 11B fehérje
BCRP	Breast Cancer Resistance Protein	Emlőrák rezisztencia fehérje
BER	Base Excision Repair	Báziskivágó javítás
BGI	Beijing Genomics Institute	Beijing Genomics intézet
BRAF	B-Raf Proto-Oncogene, Serine/Threonine Kinase	B-Raf proto-onogén, szerin/treonin kináz
BRCA1/2	Breast Cancer 1/2, Early Onset	Emlőrák 1/2, korai fehérje
BUB1	Bub1 Mitotic Checkpoint Serine/Threonine Kinase	Bub1 mitotikus ellenőrzőpont szerin/treonin kináz
BUB1B	Mitotic Checkpoint Kinase Mad3L	Mad3L mitotikus ellenőrzőpont kináz
BWT	Burrows-Wheeler transform	Burrows-Wheeler transzformáció
CCD	charge-coupled device	töltés-csatolt eszköz
CCNA2	Cyclin A2	Ciklin A2
CCNB2	Cyclin B2	Ciklin B2
CCNE1	Cyclin E1	Ciklin E1
CD79A	Cd79A Molecule	Cd79A molekula
CDC19	Minichromosome Maintenance Complex Component 2	Minikromoszóma fenntartó komplex 2 komponens

CDC20	Cell Division Cycle 20	Sejtciklus szabályozó 20
CDC25A	Cell Division Cycle 25A	Sejtciklus szabályozó 25A
CDC54	Minichromosome Maintenance Complex Component 4	Minikromoszóma fenntartó komplex 4 komponens
CDC6	Cell Division Cycle 6	Sejtciklus szabályozó 6
CDC7L1	Cell Division Cycle 7	Sejtciklus szabályozó 7
CDH1	Cadherin 1	Kadherin 1
CDK4	Cyclin Dependent Kinase 4	Ciklin-dependens kináz 4
CDKN2A	Cyclin Dependent Kinase Inhibitor 2A	Ciklin-dependens kináz inhibitor 2A
cDNS	Coding deoxyribonucleic acid	komplementer DNS
CHEK1	Checkpoint Kinase 1	Ellenőrzőpont kináz 1
CHEK2	Checkpoint Kinase 2	Ellenőrzőpont kináz 2
CHiP	Chromatin immunoprecipitation	Kromatin immunoprecipitáció
CMOS	complementary metal-oxide-semiconductor	komplementer fém-oxid félvezető
CNTRL	Centrosomal Protein 1	Centroszómális fehérje 1
CNV	Copy-number variation	Kópiaszám változás
COL1A1	Collagen Type I Alpha 1 Chain	1-es típusú kollagén alfa lánc
COSMIC	Catalogue of Somatic Mutations in Cancer	Tumorer szomatikus mutációinak katalógusa
DBF4	Activator Of S Phase Kinase	S fázis kináz aktivátor
DCC	Deleted In Colorectal Carcinoma	Vastagbél-daganatokban törlődött
ddNTP	Dideoxynucleotide	didezoxiribonukleotid
DICER1	Dicer 1, Ribonuclease Iii	Dicer 1, ribonukleáz Iii
DNAJB1	DnaJ Heat Shock Protein Family (Hsp40) Member B1	Hősokkfehérje család B1 tagja
DNS	deoxyribonucleic acid	dezoxiribonukleinsav
E2F2	E2F Transcription Factor 2	E2F transzkripció faktor
EGA	European Genome-Phenome Archive	Európai Genom-Fenom archívum
EGFR	Epidermal Growth Factor Receptor	Epidermális növekedési faktor receptor
ELF4	E74 Like Ets Transcription Factor 4	E74-szerű Ets transzkripció faktor 4
ER	Estrogen Receptor	Ösztrogén receptor
ERCC3	Ercc Excision Repair 3, Tfiif Core Complex Helicase Subunit	ERCC Kivágó javítás fehérje 3, helikáz alegység
ERK	Extracellular Signal-Regulated Kinases	Extracelluláris által jel szabályozott kináz
ESR1	Estrogen Receptor 1	Ösztrogén receptor 1

FAT4	Fat Atypical Cadherin 4	Fat atipikus kadherin 4-es tag
FDA	Food and Drug Administration	Amerikai élelmiszer és gyógyszer ügynökség
FGFR2	Fibroblast Growth Factor Receptor 2	Fibroblaszt növekedési faktor receptor 2
FLI	Flii, Actin Remodeling Protein	Flii aktin remodellező fehérje
FM-index	Ferragina-Manzini index	Ferragina-Manzini index
FOXO4	Forkhead Box O4	Forkhead Doboz O4
GATA3	Gata Binding Protein 3	Gata kötő fehérje 3
GATK	Genome Analysis Toolkit	Genom elemző programkeret
GEO	Gene Expression Omnibus	gén expressziós omnibusz
GFP	Green Fluorescent Protein	Zöld fluoreszcens fehérje
GNAS	Gnas Complex Locus	Gnas komplex lokusz
GTP	Guanosine Triphosphate	Guanozin-trifoszfát
HER2	Human Epiidermal Growth Factor Receptor 2	Humán epidermális növekedési faktor receptor 2
HIF1A	Hypoxia Inducible Factor 1 Alpha Subunit	Hipoxia által indukálható faktor 1 alfa alegység
HR	Hazard ratio	kockázati arány
ISFET	ion-sensitive field-effect transistor	Ion-szenzitív tér-hatású tranzisztor
JAK1/2/3	Janus Kinase 1/2/3	Janus Kináz 1/2/3
KAT6A	Lysine Acetyltransferase 6A	Lizin-acetiltranszferáz 6A
KDM5C	Lysine Demethylase 5C	Lizin-demetiláz 6A
KI67	Marker Of Proliferation Ki-67	Proliferációs marker
KRAS	Kirsten Rat Sarcoma Viral Oncogene Homolog	Kirsten patkány szarkóma vírus onkogén homológ
LOH	Loss Of Heterozygosity	Heterozigótaság vesztes
LRP1B	Ldl Receptor Related Protein 1B	Ldl receptor rokon fehérje 1B
MAD2L1	Mitotic spindle assembly checkpoint protein MAD2A	Mitotikus orsó összállás ellenőrző fehérje MAD2A
MAP2K4	Mitogen-Activated Protein Kinase Kinase 4	Mitogén-aktivált fehérje kináz kináz 4
MAP3K1	Mitogen-Activated Protein Kinase Kinase Kinase 1	Mitogén-aktivált fehérje kináz kináz kináz 1
MAS5	microarray suite 5.0	microarray csomag 5.0
MC1R	Melanocortin 1 Receptor	Melanokortin receptor 1
MCC	Mutated In Colorectal Cancers	Vastagbél-daganatokban mutálódó
MCM6	Minichromosome Maintenance Complex Component 6	Minikromoszóma fenntartó komplex 6 komponens

MCM7	Minichromosome Maintenance Complex Component 7	Minikromoszóma fenntartó komplex 7 komponens
MDM2	Mdm2 Proto-Oncogene	Mdm2 proto-onkogén
MDR1	Multidrug Resistance Protein 1	Multidrog rezisztencia fehérje 1
MED12	Mediator Complex Subunit 12	Mediátor komplex 12-es alegysége
MEK	Mitogen-Activated Protein Kinase	Mitogén-aktivált fehérje kináz
miRNS	Micro ribonucleic acid	mikro ribonukleinsav
MLL3	Lysine Methyltransferase 2C	Lizin metiltranszferáz 2C
mRNS	Messenger ribonucleic acid	hírvivő ribonukleinsav
MRP1	Multidrug Resistance Associated Protein 1	Multidrog rezisztencia asszociált fehérje 1
MSH2	Muts Homolog 2	Muts homológ 2
mTOR	Mammalian Target Of Rapamycin	Emlős rapamycin célpont gén
MYH11	Myosin Heavy Chain 11	Miozin nehéz lánc 11
MYH2	Myosin Heavy Chain 2	Miozin nehéz lánc 2
NF1	Neurofibromin 1	Neurofibromin 1
NGS	Next-generation sequencing	Újgenerációs szekvenálás
NIN	Ninein	Ninein
NLRP3	Nlr Family Pyrin Domain Containing 3	Nlr család pirin domént tartalmazó 3
NMF	nonnegative matrix factorization	Nem-negatív mátrix faktorizálás
NSD1	Nuclear Receptor Binding Set Domain Protein 1	Sejtmag receptor kötő Set domén fehérje 1
PAGE4	Page Family Member 4	Page család 4-es tagja
PALB2	Partner And Localizer Of Brca2	Brca2 partnere
PARP	Poly [ADP-Ribose] Polymerase 1	Poli-[ADP-ribóz] polimeráz 1
PCR	Polymerase chain reaction	Polimeráz láncreakció
PI3K	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase	Foszfatidilinozitol-4,5-biszfoszfát 3-kináz
PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha	Foszfatidilinozitol-4,5-biszfoszfát 3- kináz katalitikus alegysége
PIK3R1	Phosphoinositide-3-Kinase Regulatory Subunit 1	Foszfoinozítid-3-kináz 1 szabályozó alegysége
PLK1	Polo Like Kinase 1	Polo-szerű kináz 1
PR	Progesterone Receptor	Progeszteron receptor
PRDM16	Pr/Set Domain 16	Pr/Set domén 16
PTCH1	Patched 1	Patched 1
PTEN	Phosphatase And Tensin Homolog	Foszfátáz és tenzin homológ

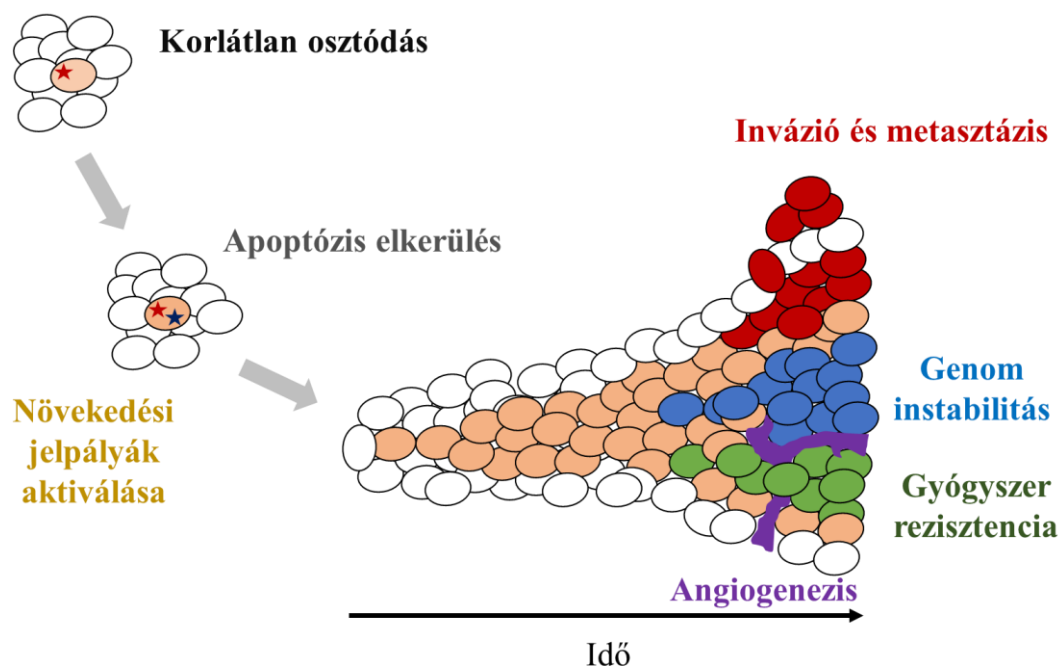
PTPN13	Protein Tyrosine Phosphatase, Non-Receptor Type 13	13-típusú receptor tirozin foszfataz fehérje
PTPN22	Protein Tyrosine Phosphatase, Non-Receptor Type 22	22-típusú receptor tirozin foszfataz fehérje
PTPRB	Protein Tyrosine Phosphatase, Receptor Type B	B-típusú receptor tirozin foszfataz fehérje
PTPRC	Protein Tyrosine Phosphatase, Receptor Type C	C-típusú receptor tirozin foszfataz fehérje
PTPRK	Protein Tyrosine Phosphatase, Receptor Type K	K-típusú receptor tirozin foszfataz fehérje
RABEP1	Rabaptin, Rab Gtpase Binding Effector Protein 1	Rab GTPáz kötő effektor fehérje 1
RB1	Retinoblastoma 1	Retinoblasztóma 1 gén
RBL1	Rb Transcriptional Corepressor Like 1	Rb transzkripció represszor-szerű 1 fehérje
RFP	Red Fluorescent Protein	Vörös fluoreszcens fehérje
RNA-seq	RNA sequencing	RNS szekvenálás
RNS	ribonucleic acid	ribonukleinsav
ROC	Receiver operating characteristics	vevő működési karakterisztika
RTK	Receptor Tyrosine Kinase	Receptor tirozin kináz
RUNX1	Runt Related Transcription Factor 1	Runt rokon 1 transzkripció faktor
SALL4	Spalt Like Transcription Factor 4	Spalt-szerű transzkripció faktor 4
SAM	Simple alignment and mapping	Egyszerű illeszkedés formátum
SDHAF2	Succinate Dehydrogenase Complex Assembly Factor 2	Szukcinát-dehidrogenáz komplex együttes faktor 2
SETD2	Set Domain Containing 2	Set domént tartalmazó 2-es fehérje
SETX	Senataxin	Senataxin
SKP2	S-Phase Kinase Associated Protein 2	S-fázis kinázhoz asszociált 2-es fehérje
SMAD2/4	Mad Homolog 2	Mad homológ 2
SMARCB1	SWI/SNF-Related Matrix-Associated Protein 1	SWI/SNF-kapcsolt mátrix-asszociált fehérje 1
SMRT	Single-molecule real time	Egy molekula, valós időben
SNP	Single nucleotide polymorphism	Egypontos nukleotid-polimorfizmus
SNV	Single nucleotide variant	Egypontos nukleotid-variáció
SPEN	Spen Family Transcriptional Repressor	Spen család transzkripció represszor
SRSF3	Serine And Arginine Rich Splicing Factor 3	Szerin és arginin gazdag splicing factor 3

SS18L1	Ss18L1, NbaF Chromatin Remodeling Complex Subunit	Ss18L1, NbaF kromatin újramodellező komplex alegység
STAT1/3	Signal Transducer And Activator Of Transcription 1/3	Transzkripció 1/3 jel továbbító és aktivátor
SUZ12	Suz12 Polycomb Repressive Complex 2 Subunit	Suz12 Polycomb represszív komplex 2-es alegység
TCF19	Transcription Factor 19	Transzkripció faktor 19
TCGA	The Cancer Genome Atlas	Tumor Genom Atlasz
TET2	Tet Methylcytosine Dioxygenase 2	Tet metilcitozin-dioxigenáz 3
TFE3	Transcription Factor Binding To Ighm Enhancer 3	Ighm enhancer kötő transzkripció faktor 3
TP53	Tumor Protein P53	p53 tumor fehérje
TP63	Tumor Protein P63	p63 tumor fehérje
TPR	Translocated Promoter Region, Nuclear Basket Protein	Transzlokált promóter régió gén
TSHR	Thyroid Stimulating Hormone Receptor	Pajzsmirigy stimuláló hormon
TTK	Ttk Protein Kinase	Ttk fehérje kináz
UBR5	Ubiquitin Protein Ligase E3 Component N-Recognin 5	Ubiquitin fehérje ligáz E3 komponens N-rekognin 5
USP6	Ubiquitin Specific Peptidase 6	Ubikvitin specifikus peptidáz 6
UV	Ultraviolet	Ultraviola
VCF	Variant call format	Mutáció formátum
VEGF	Vascular Endothelial Growth Factor	Ér endotél növekedési faktor
VHL	Von Hippel-Lindau Tumor Suppressor	Von Hippel-Lindau tumor szupresszor
WGS	Whole genome sequencing	Teljes genom szekvenálás
WHSC1	Wolf-Hirschhorn Syndrome Candidate 1	Wolf-Hirschhorn szindróma kandidáns 1
WT1	Wilms Tumor 1	Wilms Tumor 1
WXS	Whole exome sequencing	Teljes exom szekvenálás
XPO1	Exportin 1	Exportin 1
ZBTB16	Zinc Finger And Btb Domain Containing 16	Cink-ujj és BTB-domént tartalmazó fehérje 16
ZNF384	Zinc Finger Protein 384	Cink-ujj fehérje 384

2. Bevezetés

2.1. A tumorok kialakulását elősegítő molekuláris változások

A daganat kialakulása egy többlépcsős folyamat következménye, amelynek a végén egy korlátlanul osztódó sejtcsoport jön létre. A növekedés során folyamatosan felhalmozódó genetikai változások hatására specifikus fenotípusok jelenhetnek meg a tumorban, amiket a „Hallmarks of Cancer” (a „tumor védjegyei”) néven szokták összefoglalni [1]. A tumorokra jellemző fő sajátosságok az apoptózis elkerülése, a növekedési jelpályák aktiválása, a növekedés elleni szignálokra érzéketlenné válása, az angiogenezis, a korlátlan osztódás, valamint az invázió és a metasztázis képesség (1. ábra). Az első összesítés óta a jellemzők listája a genom-instabilitással, a gyulladás-serkentéssel, a sejt-anyagcsere változásokkal, illetve az immunrendszer elleni védekezéssel tovább bővült [2].



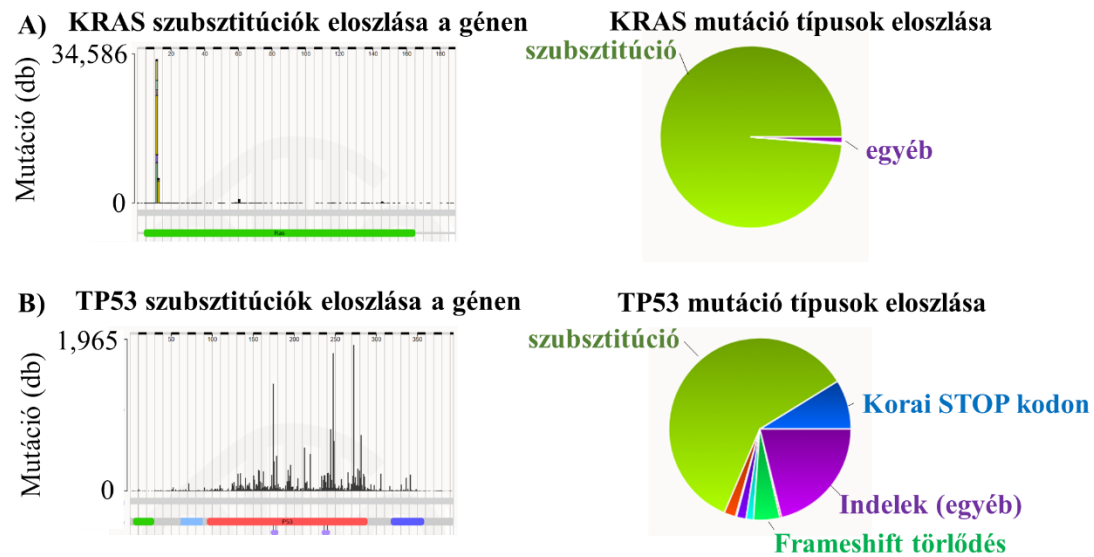
1. ábra. A tumor fejlődése során előforduló fenotípusos-változások vázlata. A korlátlanul osztódó sejtekben felgyűlő genetikai változások hatására kialakulhat a tumorban gyógyszer-rezisztencia, angiogenezis képesség, valamint metasztatizáló képesség.

Ezen jellegzetességek a növekedés során fokozatosan gyűlnek fel, a tumor kezdeti fázisaiban csak részben vannak jelen. A tumor növekedése során felhalmozott

több ezer mutáció a jelátviteli szabályozásában kulcsfontosságú géneket is érinthet. Ezen géneket onkogén illetve tumorszupresszor csoportokba osztjuk. Az onkogének támogatják a sejtosztódást, a növekedést és gátolják az apoptózist. Ezzel szemben a tumorszupresszor gének gátolják a sejtosztódást, valamint apoptózist indukálhatnak. A két géncsoport nem csak funkcióban, de a mutációk típusában is eltér.

Az onkogének serkentik a sejtosztódást, mutációik hatására általában fokozódik a működésük. Az onkogén aktiváció egyik lehetősége a kódolt fehérje mennyiségi megváltozása gén amplifikáció vagy epigenetikai módosítás következményeként. Ehhez kapcsolódó legismertebb példa az epidermális növekedési faktor receptor gén (illetve géncsalád) amplifikációja, mely az általa irányított RAS és PI3K útvonalak aktiválásához vezet [3, 4]. Az aktiváció másik lehetősége az onkogén szekvenciájának módosulása. Ezen változások általában egy (vagy néhány) aminosavat érintenek, hatásukra a fehérjék szabályozás nélkül aktiválhatják a jelátviteli útvonalakat. Erre példa a KRAS onkogén gyakori mutációja több tumor típusban. A mutációk többsége a 12. illetve a 13. aminosav pozícióban okoznak cserét a gén GTP-áz részében, ezzel konstitutívan aktiválva a fehérjét (2/A ábra) [5, 6, 7].

A tumorszupresszor génekben inaktiválós jelenségek játszódnak le a tumorokban. Ezek általában gén törlődéssel, epigenetikai csendesítéssel, vagy a kódolt fehérje aktivitását romboló mutációkkal történnek. Fontos megjegyezni, hogy a tumorszupresszor gének általában mindkét alléljának mutációja szükséges a daganat kialakulásának elősegítésére. Ennek iskolapéldája az örökletes emlő- és petefészek tumort okozó BRCA1 illetve BRCA2 génmutációk [8]. Ezen géneknél gyakran megfigyelhető a szomatikus heterozigótaság vesztes (LOH). Ilyenkor a gén vad kópiája törlődik a tumor genomából, és csak a hibásan működő kópia marad meg [9, 10, 11, 12]. A tumorszupresszor géneket érintő mutációk abban különböznek az onkogén mutációktól, hogy elszórva helyezkednek el a gén szinte bármely részén (legismertebb példa a TP53 gén; 2/B. ábra [13]) és jóval magasabb a fehérje egy részét vagy egészét törő (trunkáló) mutációk aránya (2/B ábra) [14, 15, 16]. A mutációk hatására a fehérje nem képes eredeti funkcióját ellátni.



2. ábra. A szomatikus mutációk elhelyezkedése, valamint típusának eloszlása tumor génekben [17]. A) A KRAS onkogénben a szubsztitúciós mutációk többsége egy aminosavat érintenek. B) A TP53 tumorszupresszor génben szubsztitúciók, stop kodont okozó mutációk, valamint bázis törlődések és beszúrások a gén szinte bármely részén előfordulhatnak.

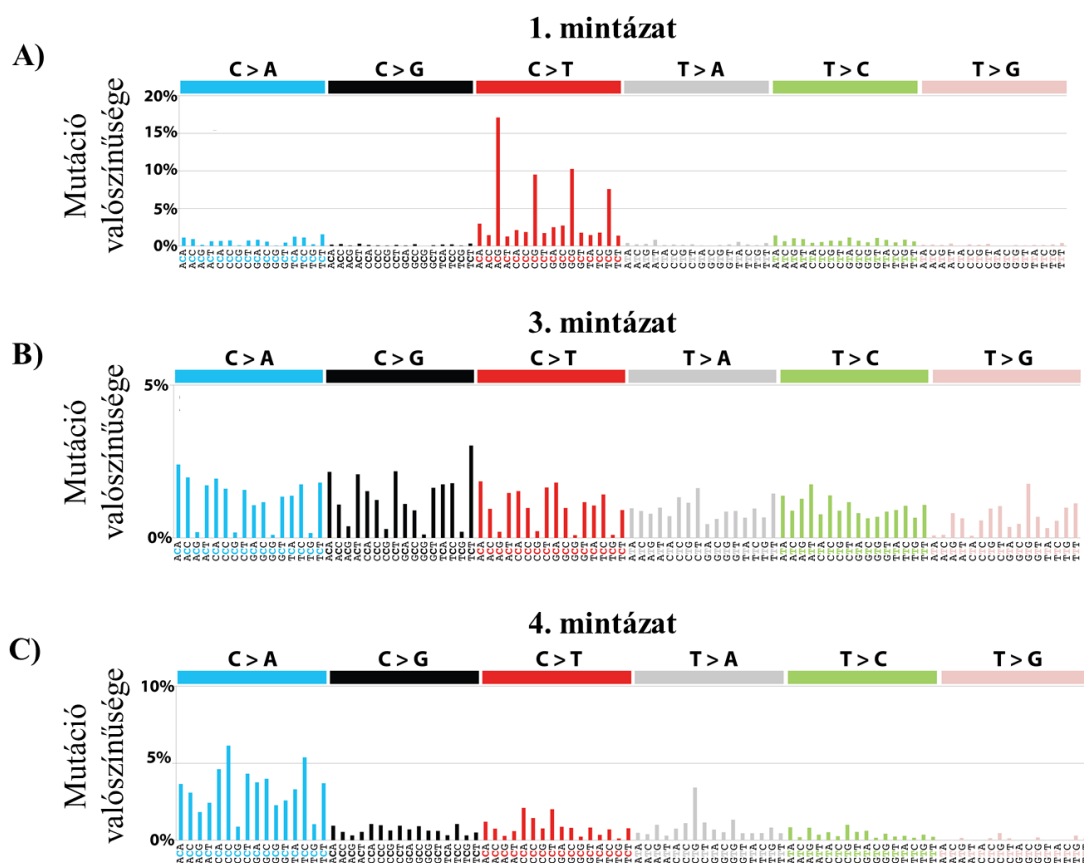
A karcinogenezis folyamata során általában nem elegendő egyetlen molekuláris változás. Nording [18] 1953-ban leírta, hogy a tumor kialakulásában fontos szerepe van a beteg korával összefüggő felhalmozott genetikai változásoknak. Knudson 1971-ben bemutatta, hogy a gyermekkori retinoblastoma kialakulásához nem elegendő az RB1 gén egyik kópiájának mutációja, szükség van a másik allélt érintő mutációra is a gén teljes inaktiválásához. Az RB1 gén funkció vesztese létrejöhet két szomatikus RB1 génmutáció vagy egy öröklött és egy szomatikus RB1 génmutáció hatására is [19].

Azóta több tumor típusban igazoltak a karcinogenezishez hozzájáruló öröklött mutációkat. Ismertebb példák a vastagbél-daganatokban az APC génben [20], az emlő és a petefészek tumorokban a BRCA1/2 génekben [8, 21], a vesedaganatokban a VHL génben [21] és a Wilms-tumorokban a WT1 génben [22] található öröklött mutációk. Mindegyik esetben gyakori a gén heterozigóta mutációja (vagy törlődése), mely mellett a tumorban szomatikus mutációval inaktiválódik a második kópia.

2.2. Tumorok mutagenezise és a mutációs mintázatok

A tumor-genomokban felhalmozott mutációk hibás sejtosztódás, endogén vagy exogén mutagének, vagy enzimes módosítások következményei. Bizonyos tumor típusok esetén ismertek a kialakulásban szerepet játszó mutagének. A két legismertebb példa a tüdőráknál a dohányzás [23], valamint a melanómáknál az UV sugárzás által okozott jellegzetes szubsztitúciók [24]. Alexandrov és munkatársai [25] a mutációk típusának gyakoriságát trinukleotid kontextusban vizsgálták. Ezekben a középső bázis cserélődött, de úgy találták, hogy fontos szerepe van a két szomszédos bázisnak is. Nem-negatív mátrix-faktorizálás segítségével 20 mintázatot azonosítottak szimulált és tumoros minták szekvencia-adataiból.

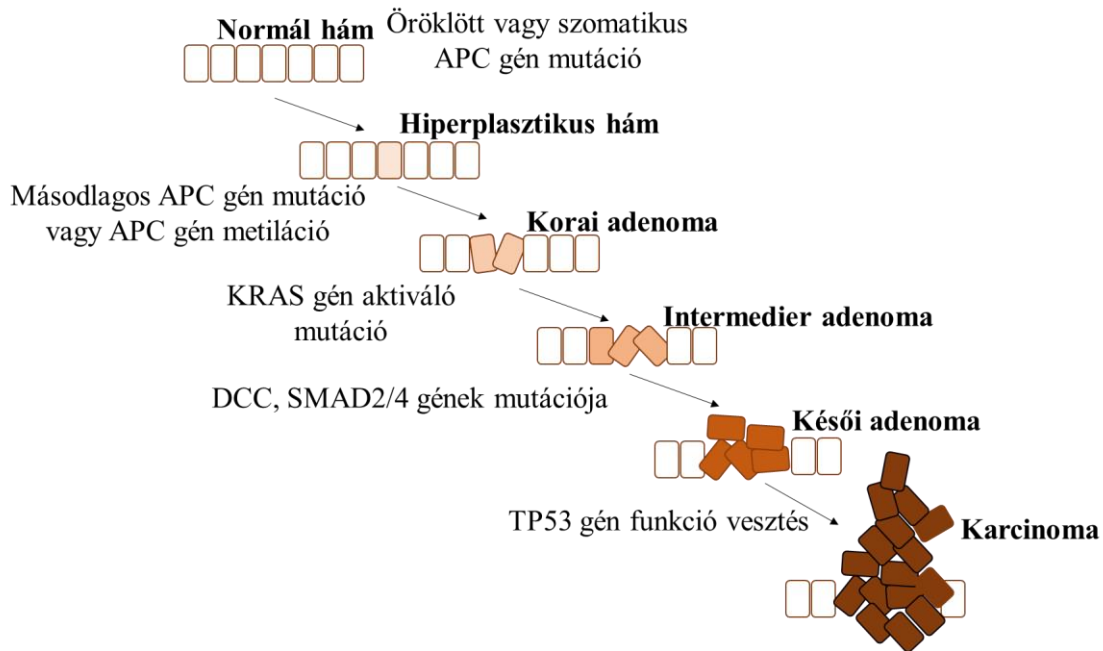
A különböző mutagén szerek és folyamatok más és más jellegű mutációkat eredményeznek. Alexandrov és munkatársai további munkájuk során [26] több tumor típusból, összesen 7,042 beteg teljes genom, valamint exom szekvenálási adatán összekapcsolták a mutációs mintázatokat genetikai és környezeti mutagénekkel. Az első mintázat (3/A ábra) a korral függött össze, a 2-es és a 13-as mintázatok az APOBEC citidin-dezamináz hibával lettek összekötve, a hármas mintázat meg a BRCA hibajavító defektushoz lett asszociálva (3/B ábra). A 4-es mintázat a dohányzással függött össze (3/B ábra), míg a 7-es mintázat az UV sugárzással.



3. ábra. A mutációs mintázatok ismertebb példái [25]. A) A beteg korával összefüggő 1-es mintázat, B) a homológ rekombináció hibával összefüggő 3-as mintázat, valamint C) a dohányzással összefüggő 4-es mintázat.

2.3. Karcinogenezis során kialakuló molekuláris változások

A karcinogenezis egyik, talán legismertebb lineáris fejlődési vonalát a vastagbél tumorok esetén írtak le [27], ezt sematikusan a 4. ábrán foglaltam össze.

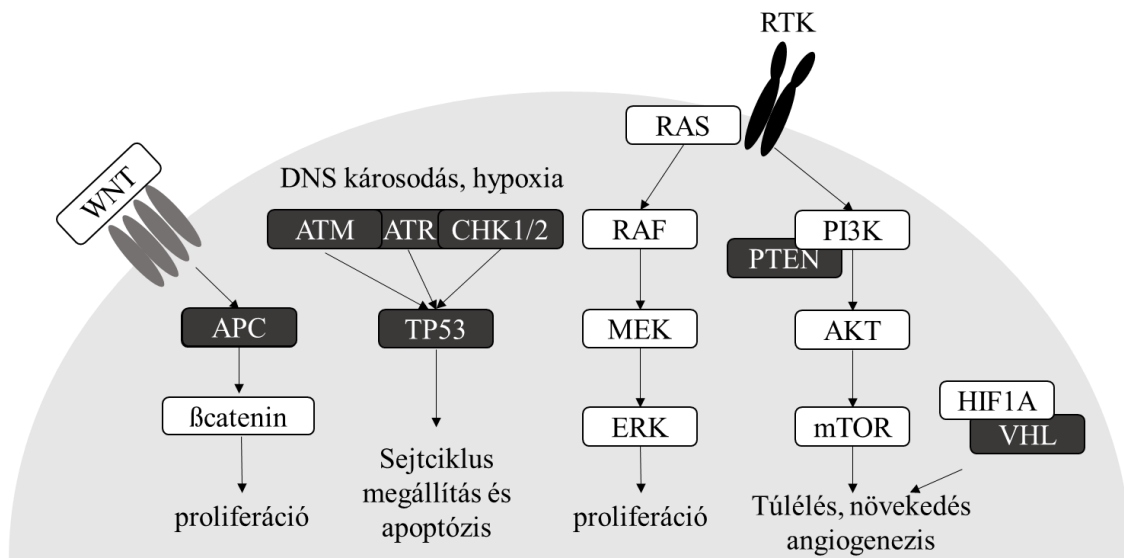


4. ábra. A vastagbél-daganatok kialakulása. Genetikai változások hatására először egy hiperplasztikus hám, majd adenoma, végül karcinóma alakul ki.

A vastagbél-daganatok számos esetben az APC gén mutációja következtében alakulnak ki. A génben elhelyezkedő változások eredményeként a normális hámból egy hiperplasztikus hám alakul ki. Emellett még az MCC génben is előfordulhatnak mutációk. A hiperplasztikus hámban kialakult elsődleges mutációk utáni metilációs változások hatására alakulhat ki a korai adenoma. A folyamat következő lépése gyakran az intermedier adenoma kialakulása a KRAS gén mutációjának hatására. Ezt követően a DCC és a SMAD2/4 gének változásainak következtében nagy- vagy késői adenoma alakulhat ki. A folyamat utolsó lépéseként gyakran a TP53 génben lévő mutáció hatására alakul ki a karcinoma [28].

Az utóbbi 20 év átfogó tanulmányaiból látható, hogy a mutációval érintett gének különbözőek az egyes tumor típusok esetében. Ráadásul az előforduló mutációk sokasága minden betegben más-más mintázatot mutat. A tumoros mutációk esetén nem egy adott gén mutációja fontos, hanem egy adott útvonal aktivitásának módosítása, ami függ a tumor szöveti eredetétől.

A tumorokban leggyakrabban érintett útvonalakat az 5. ábrán foglaltam össze. Az útvonalak közül a sok esetben a receptor tirozin kinázok (RTK), illetve az alattuk elhelyezkedő jelpályák érintettek gén változásokkal. Több tumortípusban is megfigyelték, hogy a mutációik hatására vagy a PI3K, vagy a RAS útvonalak aktiválódnak [7]. Bizonyos tumor típusoknál aktiválódhat a Wnt útvonal a β -katenin gén aktiváló mutációk [29], vagy az APC gén inaktiváló mutációk hatására [28].



5. ábra. A szolid tumorokban leggyakrabban érintett jelátviteli útvonalak (fehér: onkogén, fekete: tumorszupresszor gén).

Az emlőtumoroknál gyakori a PI3K útvonal aktiválása a PIK3CA onkogén mutáció és/vagy a PTEN tumorszupresszor gén vesztés következtében [30, 31]. A tüdő adenokarcinómákban és a vastagbél-daganatokban sokszor találhatók mutációk a KRAS, a BRAF és a PI3KCA génekben. Hatásukra mind a PI3K útvonal, mind a mitogén jelpályák is aktiválódhatnak [32, 33]. A petefészek-tumorokban gyakori mutációk találhatók a retinoblasztoma génben, valamint a sejtciklust szabályozó ciklin-dependens kinázokban [34]. A melanómákban általában a mitogén jelpályát aktiváló BRAF és NRAS géneket érintő mutációk találhatók [5]. A vastagbél-tumorokban gyakoriak a BRAF, a KRAS, a PIK3R1 és a PTEN gének mutációi [35, 36, 37]. Fontos megjegyezni, hogy ritkán fordul elő egy betegben egy jelpályán belül két aktiváló mutáció, hiszen ez nem növeli tovább a jelpálya aktivitását.

2.4. Az emlőtumörök fő molekuláris jellemzői

Az emlörák kialakulását hajlamosító genetikai tényezők között ismertek az öröklött BRCA1 és BRCA2 gének mutációi. A BRCA1 gén mutációja esetén nőkben 54% esély van az emlörák kialakulására 60 éves korig [38], míg a BRCA2 gén mutációja esetén az esély 45% körüli [39]. Öröklött BRCA1/2 génmutációval rendelkező betegeknel a tumorban gyakori a gén heterozigótaság vesztese, azaz a vad típusú allél elvesztése [10]. Fontosak még a PALB2 gén öröklött mutációi is, jelenlétük esetén 35% esély van az emlörák kialakulására 60 éves korig [40]. Ezen gének a DNS kettősszal törések homológ rekombináció általi javításában vesznek részt [41, 42, 43]

Az emlőtumörök legfontosabb molekuláris jellemzője a hormonreceptorok státuszával kapcsolatosak. Állapotuk előre jelzik bizonyos kezelések hatékonyságát [44, 45], valamint a betegek túlélésére is prognosztikusak lehetnek [46, 47, 48]. A tumorok Szt. Galleni altípusokban való sorolása az ösztrogén receptor (ER), a progeszteron receptor (PR) és a HER2 receptor (epidermális növekedési faktor receptor család) kifejeződési státusza alapján történik [49]. Három fő altípust különböztetünk meg: 1) a bazális altípusnál minden receptor negatív státuszú (tripla negatív), 2) a luminális A és B altípusokban az ösztrogén receptor státusza pozitív és 3) a HER2 pozitív altípusban a HER2 receptor státusza pozitív. A HER2 pozitív tumorok terápiája során a kemoterápiát kiegészítik HER2 gátló trastuzumab vagy pertuzumab célzott terápiával [50, 51]. A luminális tumoroknál a kemoterápia mellett gyakran alkalmaznak ciklin-dependens kináz inhibitor alapú kezelést (Palbociclib [52]). Az ösztrogén receptor pozitív betegeknel alkalmazható az anti-ösztrogén hatású tamoxifen [53]. Ösztrogén-pozitív, menopauza utáni betegeknel aromatáz gátlókat (Letrozole [54], Anastrozole [55] és Aromasin [56]), illetve az mTOR gátló Everolimus [57] kezeléseket is alkalmazhatják.

Az emlőtumorokra jellemző szomatikus mutációk találhatók a PIK3CA onkogénben, valamint a TP53 tumorszupresszor génben a betegek több mint 50%-ában [58, 59]. Számos esetben érintettek a mitogén jelpályát szabályozó MAP3K1 és MAP2K4 gének is. A luminális típusú tumorok esetén sok betegben találhatók mutációk a GATA3 transzkripciós faktor, valamint a kadherin 1 génekben. A leggyakoribb gén kópiaszám változások a PIK3CA gén amplifikációja, valamint a TP53, a MAP2K4 és az RB1 gének delécióna mindegyik altípusban. Ezeken kívül még gyakori a HER2 gén amplifikáció nem bazális szub típusokban [7].

2.5. A melanómák fő molekuláris jellemzői

A malignus melanómák kevesebb, mint 10%-ában azonosítható családi halmozódás. A familiáris melanómával diagnosztizált betegek 1/3-ában a CDKN2A génben azonosítottak mutációkat [60]. Az MC1R gén bizonyos polimorfizmusai hajlamosítanak melanómára [61], a mutációkat hordozó betegekben jelentősen megnő a diszplasztikus anyajegyek száma, valamint az UV sugárzás expozícióval arányosan fokozódik a melanóma kialakulásának esélye [62].

A sporadikus melanoma kialakulásában fontos szerepe van az UV sugárzásnak [62], valamint a növekedési jelpályákat aktiváló BRAF, KRAS, NF1 és NRAS géneket érintő szomatikus mutációknak [7]. Emellett gyakran előfordulnak a TP53, illetve a PTEN tumorszupresszor géneket érintő mutációk is [63]. Az UV sugárzás és a genetikai instabilitás miatt a melanómákban lévő mutációs teher jelentősen nagyobb a daganatok átlagához képest [64], emiatt nehéz a molekuláris célpontok pontos jellemzése.

Kemoterápiás szereken kívül célzott terápiák is alkalmazhatók melanóma eseték. A BRAF gén V600E mutációval rendelkező betegeknél a mutáns gént direkt módon, célzottan lehet gátolni Vemurafenib [65] illetve Dabrafenib [66] gyógyszerekkel. Indirekt kezelés során a BRAF gén által szabályozott mitogén jelpálya gátlására van lehetőség Trametinib [67] illetve Cobimetinib [68] gyógyszerekkel.

2.6. A petefészek-tumorok molekuláris jellemzői

A petefészek-daganatok közül a leggyakoribbak a hámeredetű szerózus daganatok, melyek differenciáltsági foka lehet alacsony (LGSC) vagy magas (HGSC). A családirag halmozott petefészek daganatos betegekben rendszerint a BRCA1 és a BRCA2 génekben lehet öröklött mutációkat azonosítani [69, 70]. A BRCA1 gén mutációja esetén 40-60% az esélye, hogy egy élet során petefészek-tumor alakuljon ki, míg a BRCA2 gén mutációja esetén ez az esély 20-35%.

A szomatikus mutációkat általában a TP53 tumorszupresszor génben azonosíthatók. A TP53 gén mutációjának prevalenciája elérheti akár a 96%-ot a magas differenciáltsági fokú szerózus tumoroknál. További szomatikus mutációk találhatók az

NF1, a BRCA1/2, az RB1, a CHEK2 és a CDK1/2 génekben [71]. A JAK1/2 és a STAT1/3 gének emelkedett kifejeződése jellemző a magas differenciáltságú tumorokban.

A petefészek-tumoroknál gyakran alkalmaznak bevacizumab célzott kezelést [72]. A kezelés a VEGF fehérjéhez kötődve lassítja az angiogenezist és a hozzá tartozó jelátviteli útvonalakat. A BRCA1/2 génmutációval rendelkező betegeknél a BER (báziskivágásos javítás hibajavító) útvonalat gátló PARP inhibitorok (pl Olaparib [73]) alkalmazhatók. Hormonterápia során általában célzott tamoxifen anti-ösztrogén terápiát, illetve a menopauza után aromatáz inhibitorokat is alkalmazhatnak.

2.7. Gyógyszeres kezelés, szelekció és gyógyszer rezisztencia

Az emlődaganatok, a petefészek-tumorok és a melanóma esetén alkalmazott kemoterápiás kezeléseket az 1. táblázatban foglaltam össze. A kemoterápiás szerek a timidilát-szintáz gátlással, a DNS szintézis gátlással, a mikrotubulus szintézis gátlással vagy stabilitásának növelésével, illetve antimetabolitként érik el hatásukat. Közös tulajdonságuk, hogy toxikusabbak a gyorsan osztódó sejtekre.

1. táblázat. Az emlődaganatok, a petefészek-daganatok és a melanomák esetén alkalmazott kemoterápiás szerek [74, 75, 76].

Kemoterápiás szer	Hatásmechanizmus	Emlő	Petefészek	Melanoma
5-fluorouracil	timidilát szintáz (TS) inhibitor	x		
Altretamine	alkiláló ágens		x	
Capecitabine	timidilát szintáz (TS) inhibitor		x	
Carboplatin	DNS szintézis gátlás	x		x
Ciklofoszfamid	alkiláló ágens		x	
Cisplatin	DNS szintézis gátlás			x
Dacarbazine	DNS szintézis gátlás			x
Docetaxel	mikrotubulus gátló	x		
Doxorubicin	DNS szintézis gátlás	x	x	
Eribulin	mikrotubulus gátló	x		
Etoposide	DNS szintézis gátlás		x	
Gemcitabine	DNS szintézis gátlás	x	x	
Ifosfamide	alkiláló ágens		x	
Irinotecan	DNS szintézis gátlás		x	
Ixabepilone	mikrotubulus stabilizáló	x		
Melphalan	alkiláló ágens		x	
Mitoxantrone	DNS szintézis gátlás	x		
Paclitaxel	mikrotubulus stabilizáló	x	x	x
Pemetreced	folát antimetabolit		x	
Temozolomide	alkiláló ágens			x
Topotecan	DNS szintézis gátlás		x	
Vinblastin	mikrotubulus gátló			x
Vinorelbine	mikrotubulus gátló	x	x	

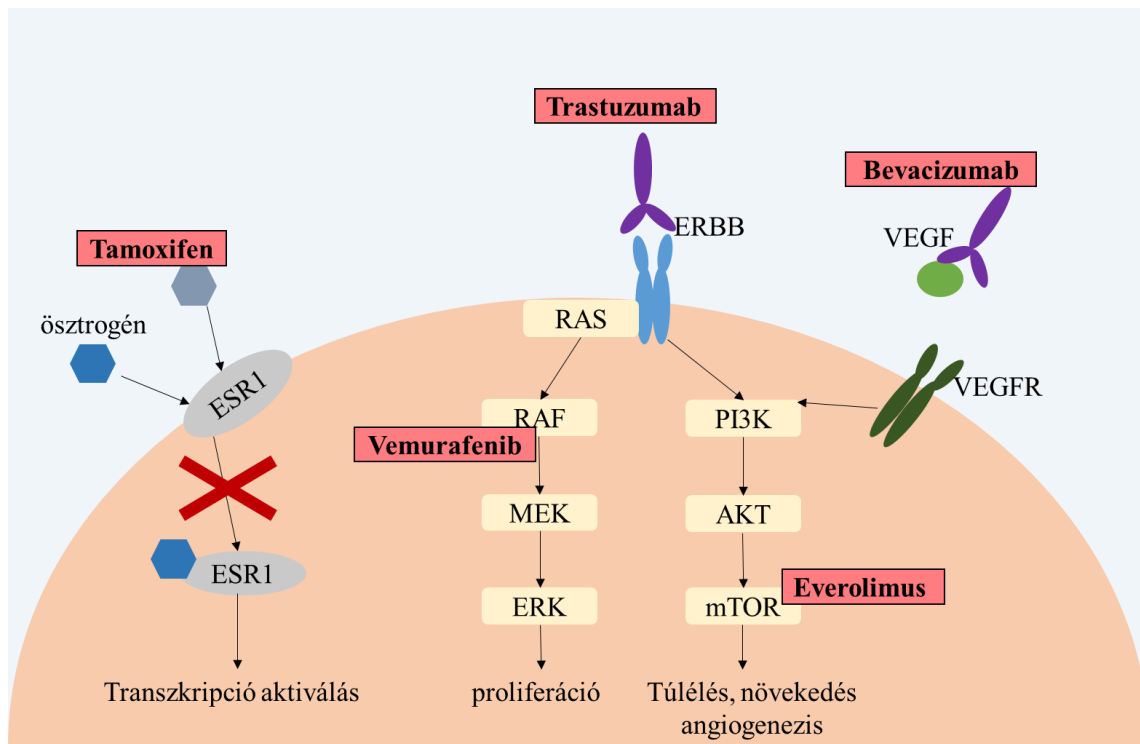
A tumorok gyógyszeres kezelése során idővel gyakran kialakul a rezisztencia mind a citotoxikus, mind a citosztatikus szerek ellen is. Rezisztenciát létrejöhet a sejt metabolizmusának megváltozása, a jelpályákat befolyásoló mutációk és epigenetikai változások hatására, vagy a célzott fehérje-módosulások miatt [77].

A rezisztencia mechanizmusok első példája a gyógyszer inaktivációja, vagy inaktív állapotban tartása. Sok kemoterápiás szer ugyanis natív állapotban inaktív, a

sejtbe kerülve több módosulási lépés után aktiválódik. Az (in)aktiváló rendszerek közé sorolható a citokróm P450 rendszer, a glutation-S-transzferáz (GST) család, valamint az UDP-glükuronil-transzferáz család. Ezen útvonalak mutációja vagy szabályozásuk módosulása általában csökkenti a terápiák hatékonyságát és növeli a sejtek rezisztenciáját.

A rezisztencia egyik gyakran vizsgált mechanizmusa a gyógyszer efflux, mely során a sejtek kipumpálják a hatóanyagokat. Ismertebb példa a sejtek homeosztázisában fontos szerepű ABC transzporter család. Ezen fehérjék nukleotid kötő, valamint variábilis transzmembrán részekből tevődnek össze. Amikor a megfelelő szubsztrát a transzmembrán részhez kötődik, a nukleotid kötő részen egy ATP molekula hidrolízise hatására konformáció változáson megy keresztül a transzmembrán rész, kipumpálva a sejtől a szubsztrátot. A tumoros gyógyszer-rezisztenciával három fő transzporter fehérje-típust hoztak összefüggésbe: a multidrog rezisztencia fehérje 1-et, a multidrog rezisztencia asszociált fehérje 1-et, és az emlő tumor rezisztencia fehérjét.

Az emlőtumороk, a petefészek tumorok és a melanóma esetén több célzott terápiára van lehetőség, ezeket a 6. ábrán foglaltam össze. Az ösztrogén-pozitív tumoroknál tamoxifen kezeléssel lehet gátolni az ösztrogén receptort. A HER2 pozitív tumoroknál trastuzumab monoklonális antitesttel lehet célolni a receptort. Bevacizumab kezeléssel lehet gátolni az érújdonképződést, így csökken a tumor vérellátása. Aktív PI3K jelátviteli útvonal esetén alkalmazható az mTOR gátló everolimus. A BRAF gén V600E mutációval rendelkező tumoroknál alkalmazható a vemurafenib, amely célzottan gátolja a mutációval rendelkező fehérjét.



6. ábra. Az emlőtumorok, a petefészek-daganatok és a melanoma esetén alkalmazott fő célzott terápiás szerek.

A célzott terápia előrehaladtával gyakran jelennek meg rezisztenciát okozó mutációk, amelyek hatására a tumor nem fog reagálni az alkalmazott gyógyszerre. Némkissejtes tüdőráknál megfigyelték, hogy a gefitinib valamint az erlotinib kezelése után rövid idővel rezisztencia alakul ki új EGFR génmutációktól [78, 79]. A célzott terápia elleni rezisztencia kialakulásának másik lehetősége egy jelpályán lévő, célzott fehérje alatti mutáció. Ismert példák a KRAS, a BRAF, valamint a PIK3CA gének mutációja, esetükben a receptor tirozin kináztól függetlenül aktiválják a növekedési jelpályákat [80, 81, 82], így nem alkalmazhatók az RTK ellenes terápiák.

A célzott terápiás szerek elleni rezisztencia kialakulhat a jelátviteli útvonalak „váltásával”. A tamoxifen elleni rezisztencia általában az ösztrogén receptor megváltozott kifejeződési szintje hatására alakul ki [83].

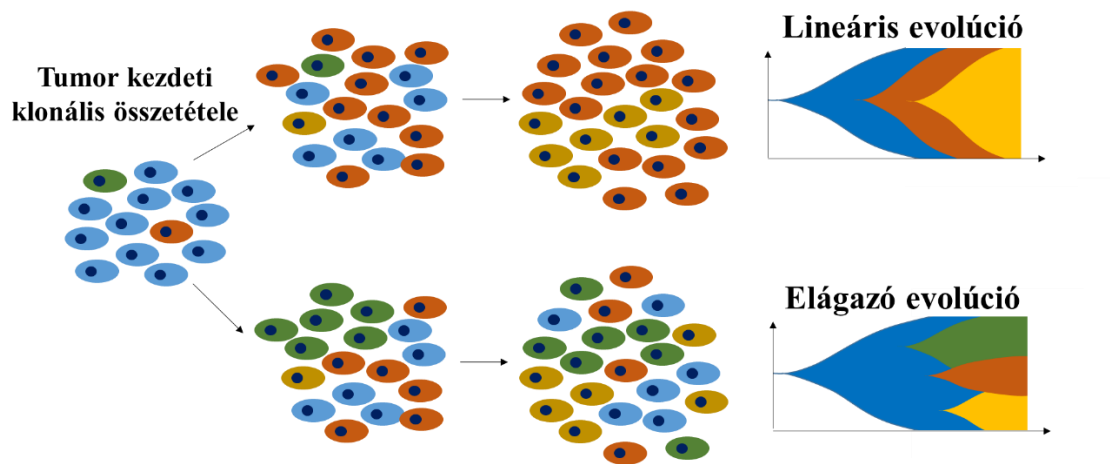
2.8. A tumor evolúció modelljei

2.8.1. A klonális szelekció

Az 1979-ben leírt klonális evolúció modellje szerint a tumor egy mutáns sejtől indul ki, osztódása és növekedése során új mutációkat halmoz fel [84]. Az új mutációk hatására új szubpopulációk (vagy szubklónok) jelennek meg, ezek tovább növekednek és osztódnak [85]. A szerzett mutációk jelentős részének nincs, vagy csak minimális hatása van a sejtek fenotípusára, ezeket *utas* („passenger”) mutációknak nevezzük. A mutációk kisebb része tartozik az ún. *vezető* („driver”) mutációk kategóriájába, ezek erősen befolyásolhatják a sejtek fitneszt. A *vezető* mutációk növekedési előnybe részesíthetnek egyes sejtcsoportokat, eredményül képesek lesznek idővel túlnőni a tumor sejtpopulációban. A tumor evolúciónak két fő modellje van: a lineáris és az elágazó klonális evolúció [86].

A lineáris evolúció során a megjelenő új klónok szinte teljesen túlsúlyba kerülnek a daganatban (7. ábra). Ennek eredménye, hogy egy adott időpontban a tumor nagy része általában egy klónból tevődik össze.

Az elágazó evolúció során a tumorban több klón és szubklón fejlődik párhuzamosan. A klónok szinte egyensúlyban élnek egymás mellett, nem nyomják el egymást, és közel azonos az osztódási sebességük (7. ábra). A modell hátránya, hogy szinte elképzelhetetlen, hogy teljesen egyensúlyban létezzenek az egyes klónok. Bizonyos genetikai és környezeti változások (beleértve a gyógyszeres kezelést) hatására az egyensúly idővel eltolódik.



7. ábra. A lineáris és az elágazó tumor evolúciós modellek. A lineáris evolúció során egy időpontban egy domináns klón van jelen. Az elágazó evolúció esetén több klón lehet párhuzamosan jelen a tumorban.

Mindkét elmélet alapján a tumor genetikailag folyamatosan változik, a legnagyobb változást a *vezető* mutációk okozzák. Egy tumor növekedése során feltehetően a két modell „ötvöze” jelentkezik, tehát a tumor klónok egyensúlyban élnek egymással, viszont némely változások hatására egyes klónok képesek kiszorulni, vagy túlnőni a többiekhez képest.

2.8.2. A tumor összejt modell

A tumor összejt modell szerint nem minden tumor sejtből képes új tumor nőni, vagy fennmaradni, erre csak néhány sejt képes. A tumor összejtek osztódhatnak újabb összejtekre, vagy nem-tumorigén sejtekre. A modell szerint a tumorokban megfigyelt heterogenitás az összejtek klonális fejlődéséből származik, ahol maga az összejtekben halmozódnak fel a mutációk. A tumor összejteket már több tumor típusban is megfigyelték, mint például a leukémiában, a glioblasztómában, az emlőtumorokban és a prosztata daganatokban.

2.9. Az intra-tumor heterogenitás és vizsgálata

A tumorok genetikai állománya időben folyamatosan változik, növekedés során új mutációk jelennek meg. Ha egy mutáció hatására javul a tumorsejt osztódási képessége, akkor az képes lesz túlnőni elődeit, ezzel új klónt létrehozva. A folyamat eredménye,

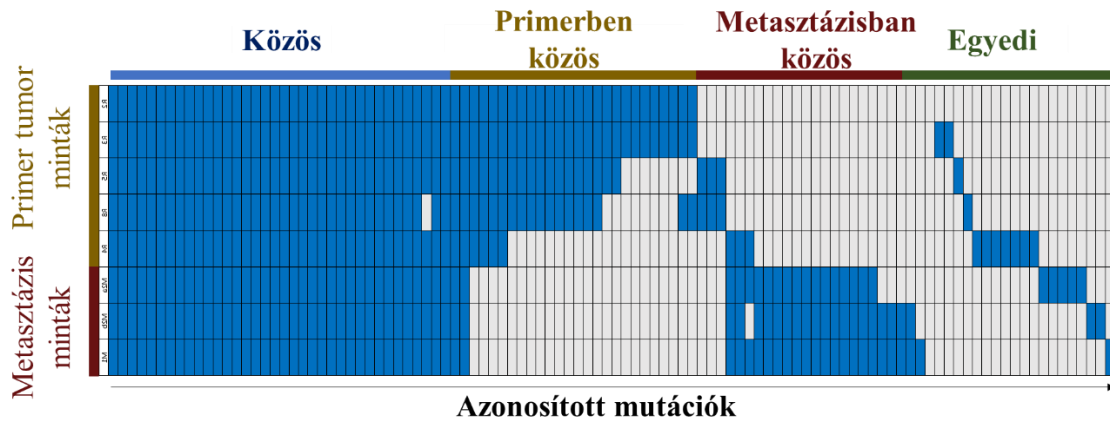
hogy egy tumoron belül sok mutáció halmozódik fel, a tumorsejt-populációban különböző arányokban lesznek jelen.

Az utóbbi években több intra-tumor heterogenitással foglalkozó vizsgálat jelent [87, 88, 89, 90]. Ezek két fő csoportra bonthatók. Az első csoport az egy tumoron belüli térbeli heterogenitást vizsgálja több régióból vett minták összehasonlításával. A második csoport a tumor heterogenitás időbeli lefolyását követi, sok esetben a kezelések hatását vizsgálva a genetikai összetételre, amit longitudinális szekvenálásnak nevezünk.

2.10. Multi-régió szekvenálást alkalmazó vizsgálatok

Multi-régió szekvenálás során egy tumor több részéből izolálunk DNS-t, illetve külön mintaként szekvenálunk. Segítségével lehetőség van megvizsgálni egy tumoron belüli heterogenitást, speciális esetekben lehetőség van a primer tumort és a metasztázist összehasonlítani. Ezen mérések előnye az egymintás szekvenáláshoz képest, hogy az azonosított mutációk felhasználásával filogenetikai elemzésekre is van lehetőség, képet adva a tumor evolúciójáról, valamint a heterogenitásának mértékéről. Multi-régió szekvenálást alkalmazó tanulmányok világossejtes vesetumороk [87, 88], petefészek tumorok [89] és tüdő adenokarcinómák [90] esetén is megjelentek.

Az egyik első multi-régió szekvenálást világossejtes vesetumoros betegeken alkalmazták [88]. A minta gyűjtés során a primer tumorokból, valamint a metasztázisokból is izoláltak és szekvenáltak DNS-t. Az exom szekvenálásból végzett szomatikus mutáció keresés után az egyes primer tumor régiókban, valamint a metasztázisokban azonosított mutációkat egy mátrix segítségével hasonlították össze (8. ábra).



8. ábra. Az egyes vesetumor részletekben (régiókban) azonosított (kék) és hiányzó (szürke) mutációk. A sorok a régiókat, az oszlopok a mutációkat jelölik.

A mutációk egy része csak a primer tumorra voltak jellemzők, a metasztázisban nem voltak azonosíthatók. Ennek lehetséges (és valószínű) magyarázata, hogy az áttétet okozó klón vagy klónok a primer tumor egy korábbi evolúciós időpontjából származnak. Az ábrán látható, hogy a metasztázisok esetén is sok volt a közös mutáció, egy részük viszont nem volt azonosítható a primer tumorban. Egy lehetséges magyarázat, hogy a primer tumor egy olyan klónjából származott az áttét, amiben már jelen voltak az egyedi mutációk, továbbá a metasztázis helyén való növekedés során új mutációk halmozódtak fel.

Az emlőtumoroknál hasonló elágazó evolúciós mintázatot azonosítottak [91]. Úgy találták, hogy az áttétet okozó klónokat a primer tumorban is lehetett azonosítani, bár ez nem minden esetben volt lehetséges. A tumor-heterogenitás mértéke függött a tumor méretétől is.

A petefészek-daganatok multi-régió szekvenálás és SNP (egy pontos nukleotid-polimorfizmus) hibridizáció alapú mérések felhasználásával megfigyelhető volt a tumorokban az elágazó evolúció [92, 93]. A magas szintű klonális expanszió magasabb heterogenitással, illetve rosszabb prognózissal járt. A betegekben két inváziós topológiai modellt azonosítottak: 1) a csillag alakú topológia, ahol minden áttét a primer tumorból indult, valamint 2) az elágazó topológia, ahol bizonyos metasztázis sejtek korábbi áttétekből származtak.

2.11. A DNS szekvenálás

A DNS szekvenálás során egy DNS molekulában található négyféle bázis sorrendjét határozzuk meg. A klinikai diagnosztikában jelenleg is felhasznált módszer a Sanger-féle szekvenáláson alapszik. A Sanger szekvenáló eszközök rendkívül pontosak, viszont hátrányuk, hogy mivel az egyes DNS szakaszokat külön kell vizsgálni a kapilláris elektroforézis során, emiatt alacsony az áteresztőképességük. Az utóbbi években a molekuláris diagnosztikai fejlesztések a „nem kapilláris elektroforézis” alapú módszerek felé fordultak. Ezen módszerek előnye, hogy több millió DNS fragmens párhuzamos vizsgálatára adnak lehetőséget, viszont pontatlanabb az eredményük.

2.11.1. Sanger szekvenálás

A Sanger szekvenálás a polimeráz láncreakcióhoz (PCR) hasonlóan a komplementer szál szintézisen alapul. A Sanger szekvenálás során a megszokott nukleotidokon kívül kis mennyiségű ddNTP-t (didezoxiribonukleotid) is használnak [94]. Ezen nukleotidokról hiányoznak a pentóz váz 3' szénatomjairól az oxigéncsoportok, beépülésük során irreverzibilisen gátolják a polimerizációs folyamatot. A reakció terméke több, különböző hosszúságú DNS darab lesz.

Az első Sanger szekvenálók négy párhuzamos reakcióban végezték a szekvenálást, minden reakcióhoz egy specifikus ddNTP volt található. A polimeráz láncreakció, valamint a négy minta poliakrilamid gélelektroforézissel végzett párhuzamos elválasztása után lett leolvasható a szekvencia. Manapság az egész reakció egy térben történik, ahol mind a négy ddNTP nukleotidon más színű fluoreszcens molekula található. A fragmensek méret szerinti elválasztása kapilláris elektroforézissel történik. A kapilláris végén lévő detektor felhasználásával lehet azonosítani az elválasztás során aktuálisan áthaladó fluoreszcens ddNTP bázisokat tartalmazó DNS fragmenseket [95]. Jelenleg a Sanger szekvenálással akár 384 mintát vizsgálhatunk párhuzamosan, a fragmensek hossza elérheti az ezer bázispárt is.

A sokéves fejlesztéssel mára sikerült elérni a 99,99%-os szekvenálási pontosságot, emiatt a Sanger szekvenálókat továbbra is gyakran alkalmazzák az onkológiai diagnosztikában. Például, több tumortípusnál Sanger szekvenálással állapítják meg az RTK gátló terápiát kizáró KRAS és BRAF génekben a mutációkat.

2.11.2. Újgenerációs szekvenálási módszerek

Az újgenerációs szekvenálás két típusát különböztetjük meg: 1) az amplifikáción alapuló újgenerációs szekvenálás (második generációs szekvenálás), és 2) az egy-molekula szekvenálásán alapuló (harmadik generációs szekvenálás) technikák.

A DNS amplifikációján alapuló újgenerációs szekvenálás

Illumina

Az Illumina amplifikációs módszerét híd-amplifikációnak (bridge-amplification) nevezik. A minta DNS-t összetörik, a fragmentumok mindkét végére adapter szekvenciákat kötnek, majd a DNS kettősszálakat denaturálják. Az előkészített DNS fragmensek az áramlási cellára (angolul *flowcell*) adagolva képesek lesznek hibridizálni a cella felületén lévő oligonukleotid adapterekhez. Az amplifikáció során először elhajlítják a DNS szálakat, így a szabad végük hibridizálni fog az áramlási cella felületén lévő másik lehorgonyzott adapterekhez. A második lépés a komplementer szál szintézise, ahol primerként a lehorgonyzott adapterek vannak felhasználva. A komplementer szálak szintézise után a kettős szálakat újból denaturálják. Az amplifikáció többszöri ismétlés után az áramlási cella egy területén ugyanannak a fragmentumnak sok másolata lesz megtalálható, amit klaszternek neveznek. Szekvenálás előtt eltávolítják a klaszterekben lévő komplementer irányú szálakat. A szekvenálási reakció során a szekvenáló készülék fluoreszcensen jelölt dNTP-eket adagol az áramlási cellára. Ezen nukleotidokat a polimerázok be tudják építeni a komplementer szálakba, azonban a fluoreszcens jelölések megakadályozzák a komplementer szálak további szintézisét. A következő lépésben gerjesztik a fluoreszcens festékeket, amiket egy CCD kamera segítségével detektálnak [96]. Végül eltávolítják a szintézist gátló fluoreszcens jelöléseket, így a következő fluoreszcens nukleotidok is képesek lesznek beépülni.

Ion Torrent

Az Ion Torrent technológiája manapság az egyik legújabb és leggyorsabban terjedő szekvenálási módszer a piacon. A módszer előnye, hogy elődjeikkel szemben nem alkalmaz fénydetektálást a szekvenálás követésére, továbbá nincs szüksége speciálisan kifejlesztett nukleotidokra. Az Ion Torrent minta előkészítése emulziós PCR (emPCR) reakcióval

történik. Az emPCR során egy víz/olaj emulziót hoznak létre, minden vízcseppbe egy gyöngy, egy polimeráz, a dNTP-k, és egy DNS szál kerül be. A gyöngy felszínén találhatók a lehorgonyzott primerek, amiről a komplementer szálak szintézise megy végbe. A PCR reakció végén minden gyöngyön egy felsokszorozott DNS lesz található. A gyöngyök egy speciális, miniatűr üreget tartalmazó ún. „ion chip”-re kerülnek, ahol minden üregbe egy gyöngy fér el.

Az Ion Torrent szekvenálás alapja a DNS szintézise közben felszabaduló hidrogén ion detektálása az ún. „ion chip” félvezető érzékelővel. A technológia CMOS-kompatibilis (ez egy félvezető építési technológia), ami jelentősen csökkenti az eszköz költségét. A csip három mikrométeres üreget tartalmaz, az üregek alját egy fénoxid érzékelő réteg alkotja. Az üregben a nukleotidok beépülésekor felszabaduló proton pH-változást okoz, ez az érzékelő rétegben felületi potenciál-változásként jelenik meg. A potenciál-változás a réteg alatti ionérzékeny tranzisztortban (ISFET, azaz ion-sensitive field-effect transistor) feszültségváltozásként mutatkozik [97].

Az egy-molekulás újgenerációs szekvenálás

Pacific Biosciences

Ezt az eljárást angolul *single-molecule, real time* szekvenálásnak (SMRT) nevezik, amit „egymolekula, valós idejű” szekvenálásnak fordíthatunk. A szekvenálást egy üreges lap segítségével végzik, mindegyik üreg tartalmaz egy, a lapra rögzített polimeráz enzimet, amit folyamatosan gerjesztenek. Szekvenálás során pirofoszfát csoporton fluoreszcens festékkel jelölt nukleotidokat öntenek a lapra, amiket a polimerázok a komplementer szál szintézisére használnak fel. A beépülésük ideje alatt a készülék le tudja mérni a nukleotidok pirofoszfát csoportjára rögzített fluoreszcens festékeket [98]. A módszer hosszú leolvasásokat képes generálni, viszont hátránya, hogy a folytonos gerjesztés miatt a polimeráz enzimek egy idő után destabilizálódnak, ami szintézis hibákhoz vezet.

Oxford Nanopore

Az Oxford Nanopore által fejlesztett szekvenáló készülék az elődjétől teljesen eltérően nem a komplementer szál szintézisére alapszik. A nanopórusos szekvenálás a négy bázis különböző „elektromos” tulajdonságaira épült. A szekvenálás során az egyszálú DNS darabokat 1 nm átmérőjű nanopórusokon vezetik át. Az egyes nukleotidok

áthaladásakor különböző mértékben változtatják meg a pórusokon átmenő áramerősséget. Nincs szükség a DNS sokszorosítására, minden egyes szál egyenként kerül leolvasásra. Jelenleg 2D szekvenálást végeznek, ahol egy fragmens elsődleges szekvenálása után a komplementer szál szekvenálása következik [99].

2.12. Orvosi kutatásban alkalmazott újgenerációs szekvenálási technikák

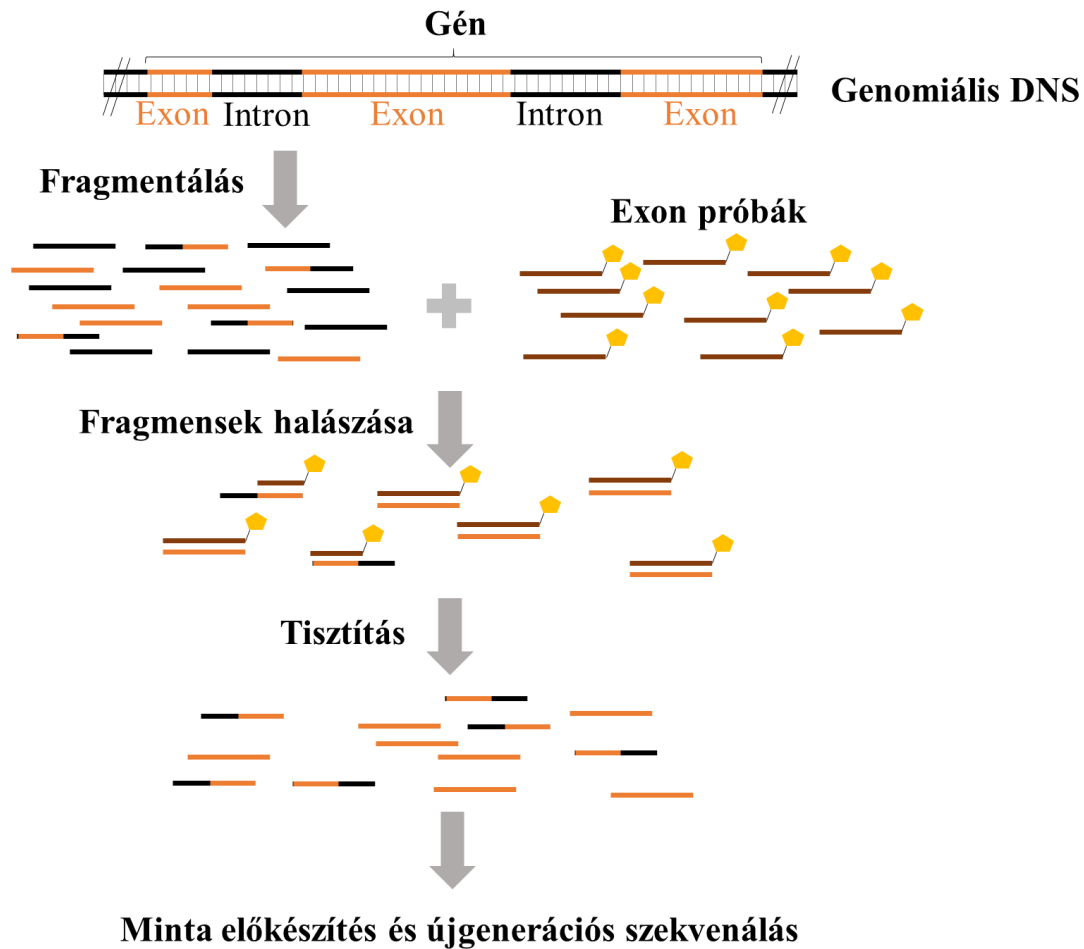
Az újgenerációs szekvenálás fő felhasználási területeit a 2. táblázat foglalja össze.

2. táblázat. Újgenerációs szekvenálás fő alkalmazási területei.

Módszer	Cél
Teljes genom	Új genom összeállítása
	Polimorfizmusok, mutációk, keresése; Strukturális variációk azonosítása [100]
Teljes exom szekvenálás	Genom kódoló régióiban való polimorfizmusok és mutációk azonosítása [101]
Célzott szekvenálás	Cél gének vizsgálata (1-1000) [102]
Transzkriptom szekvenálás	Génkifejeződés kvantifikálás [103]
	MikroRNS génkifejeződés vizsgálata [104]
Epigenetikai szekvenálás	Fehérje-DNS interakciók azonosítása [105]
	Metilációs változások azonosítása [106]
Metagenomika	Mikroorganizmus összetétel meghatározás [107]

Az újgenerációs szekvenálásnak az eredeti felhasználási területe a teljes genom szekvenálás (WGS, whole genome sequencing), ahol egy egész genomot párhuzamosan vizsgálnak. *De novo* genom szekvenálás esetén a vizsgált organizmus genom bázissorrendje általában még ismeretlen, a mérés adataiból egy elsődleges genom vázlat összeállítására van lehetőség. Manapság egyre gyakrabban végeznek teljes genom újra szekvenálást betegséggel kapcsolatos gént nem kódoló mutációk és nagyobb strukturális változások azonosításának céljából [108].

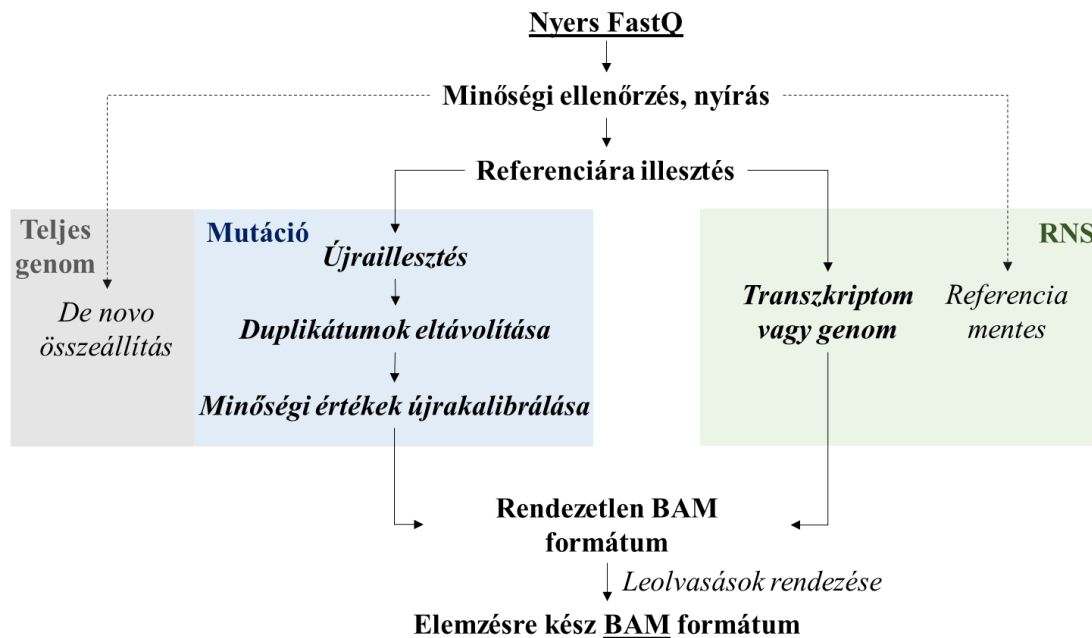
A teljes exom szekvenálás a genomnak fehérjét és RNS-t kódoló szakaszainak a vizsgálatára ad lehetőséget. Alapja, hogy a szekvenálás előtt a darabolt genomiális DNS-ből megfelelő próbák segítségével kihalásszuk az érdekelt szakaszokat (9. ábra). A technika előnye, hogy csökkenti a mérés költségeit, viszont a nem-kódoló szakaszokról nem ad információt. A teljes exom szekvenálás egyik változata a célzott szekvenálás, ahol csak 10-1000 gén (vagy genom szakasz) párhuzamos vizsgálata a cél.



9. ábra. Teljes exom szekvenálás során alkalmazott DNS előkészítés lépései. A vizsgálandó DNS-t tördelik, majd exon próbák felhasználásával kihalásszák a gént kódoló szakaszokat. A kihalászott szakaszokat végül tisztítják és előkészítik az újgenerációs szekvenáláshoz.

Transzkriptom illetve microRNS szekvenálás során a sejtekben lévő összes RNS mennyiségét és minőségét analizálják. Az izolált RNS-t először a végén lévő poli-A farok segítségével tisztítják. Ez után az RNS-t töredezik, majd a fragmenseket véletlenszerű bázissorrendű primer keverékkel visszafordítják c-DNS-é (komplementer DNS). Ezen mérések segítségével lehetséges az egyes gének kifejeződési szintjét mérni, valamint új géneket és gén izoformákat azonosítani. Több minta esetén lehetséges a differenciális génkifejeződés elemzés, amivel össze lehet hasonlítani a gének kifejeződési szintjét két vagy több minta-csoport között (pl. egy gyógyszerrel kezelt és kezeletlen csoportok között).

(genom elemző programcsomag) [111] ajánlásából származó nyers FastQ file elemzési lépéseit a 11. ábrán mutatom be. Három lehetséges feldolgozási lehetőség van attól függően, hogy mutációs, génkifejeződési, vagy epigenetikai elemzéseket végzünk. Mindegyik esetben eredményül egy közös formátumú BAM file az eredmény.



11. ábra. Nyers FastQ fájlok feldolgozási lépései. A leolvasások nyírását és minőségi ellenőrzését követően illesztjük egy referencia genomra (RNS szekvenálás esetén akár transzkriptomra), majd rendezzük.

2.13.3. Adatok minőségének ellenőrzése és nyírás

A nyers adatok feldolgozásának első lépése a minőségi ellenőrzés. A legismertebb ellenőrző szoftver a minőségi paraméterekről grafikus megjelenítést is készítő *FastQC* (<https://www.bioinformatics.babraham.ac.uk>) csomag. A legfontosabb minőségi paraméterek a leolvasások száma, a leolvasások átlagos minősége, a leolvasott bázisok pozíció szerinti minőségi eloszlása, a GC bázisok aránya, valamint a minta előkészítés során keletkezett szennyeződések (pl. adapterek, primerek) mennyisége. Azon esetekben, ahol egy vagy több paraméter nem megfelelő, ott érdemes az adatokat nyírni (angolul *trimming*).

Az adatok nyírása alatt a nem megfelelő leolvasások, vagy leolvasás végek eltávolítását értjük. Ezt több szoftverrel végezhetjük, legismertebb példák a *trimmomatic*

[112] és a http://hannonlab.cshl.edu/fastx_toolkit/contact.html honlapon elérhető *FastX-toolkit*. Ezen szoftverek segítségével képesek vagyunk eltávolítani a rossz minőségű leolvasásokat és leolvasás részeket, valamint levágni a kontaminációk körüli leolvasási végeket. Az adatok nyírása után érdemes a minőségi ellenőrzést újra elvégezni. Ha a minőségi problémák továbbra is jelen vannak, akkor szigorúbb kritériumokkal szükséges tovább nyírni az adatokat.

2.13.4. A leolvasások illesztése

A szekvenciák illesztése során kikeressük a leolvasások (vagy szekvencia) pozícióját egy referencián, valamint a leolvasások egyes bázisait a lehető legjobban megfeleltetjük a referencia szekvenciával. Bár a feladat egyszerűnek tűnhet, az újgenerációs szekvenálási adatok illesztését két faktor nehezíti: 1) a referencia szekvencia általában egy teljes genom szekvenciája, ami a Humán Genom esetén több, mint 3 milliárd bázisból tevődik össze, valamint 2) egy mérés alatt 10-100 millió leolvasást mérünk párhuzamosan. A nagy adat mennyiség miatt a klasszikus *Smith-Waterman* [113] vagy a nagyságrendekkel gyorsabb *BLAST* [114] alapú illesztőkkel végzett kiértékelés hetekig tartana egy átlagos asztali számítógépen.

Az utóbbi években megjelentek olyan illesztő algoritmusok, melyek az újgenerációs szekvenálási adatok gyors feldolgozását teszik lehetővé. Ezen szoftverek alapja, hogy a referencia szekvenciáról elkészített *Burrows-Wheeler* transzformált (BWT) és *Ferragina-Manzini* index (FM-index) felhasználásával lineáris időben lehet keresni nagy genomokban, ezzel felgyorsítva az illesztés folyamatát [115]. Az illesztés során úgynevezett *heurisztikákat* (egyszerűsítéseket) is alkalmaznak a feldolgozáshoz szükséges idő további csökkentésére.

2.13.5. Mutáció és epigenetikai alapú szekvenálási adatok feldolgozása

Az legismertebb illesztő programok a mutációs és az epigenetikai adatok kezelésére a *BWA* [116], a *Bowtie2* [117] és a *Novoalign* (<http://www.novocraft.com/products/novoalign/>) szoftverek. Az illesztés után az adatokat SAM (Simple Alignment and Mapping) formátumról a tömörített BAM (bináris SAM) formátumra alakítjuk át, majd rendezzük az illesztett leolvasásokat a kromoszóma, valamint kromoszómán belüli pozíció alapján. Mivel az illesztés során egyszerűsítéseket

alkalmazunk a gyors feldolgozás érdekében, így nem minden esetben kaphatjuk a legjobb eredményt. A pontatlanságok főleg a nagyobb méretű *indelek* („inszerciók és deléciók”) környezetében fordulhatnak elő, ahol a szoftverek inkább több szubsztitúciós hibával illesztenek, ezzel kihagyva a kérdéses indeleket. Emiatt mindig érdemes az illesztett leolvasásokat újrailleszteni a hibák kijavítása érdekében. Újraillesztés során a szoftver először kikeresi azon régiókat, ahol sok egymáshoz közeli szubsztitúció található, majd ezen régiókban kisebb vagy nagyobb indeleket bevezetve megpróbálja csökkenteni a szubsztitúciók számát [111].

Az újraillesztés után a duplikátum leolvasások kezelése az utolsó lépés. Ilyenkor egy programmal megjelöljük az ismétlődő leolvasásokat, hogy csökkentsük a minta előkészítés során keletkezett műtermék leolvasások számát, illetve a zaj arányát.

2.13.6. Az RNS szekvenálási adatok alap feldolgozása

Az RNS szekvenálási adatok feldolgozása során többféle információt tudunk kigyűjteni, attól függően, hogy teljes genomra, vagy csak a transzkriptomra illesztjük a leolvasásokat [118].

A transzkriptomra illesztés manapság egyszerű és gyors feladat. Az elemzés során az ismert gének, valamint gén izoformák szekvenciáját használjuk referenciaként. A módszer hátránya, hogy nincs lehetőség új gének keresésére, csak az ismert gének kifejeződési szintjéről nyerünk információt.

Teljes genomra történő illesztés során a teljes genomot használjuk referenciaként, így a bonyolultabb megközelítésű *spliced alignment* segítségével tudunk pontosan illeszteni. Ilyen esetben a leolvasások gyakran egy gén két különböző exonjára illeszkednek, amik egy intronnal vannak elválasztva. A módszer hátránya, hogy a folyamat jóval lassabb, és nagyobb a hibalehetősége. Előnye, hogy nem csak az ismert génekről, valamint gén izoformákról kapunk információt, hanem képesek vagyunk új géneket és izoformákat is azonosítani. Manapság leggyakrabban használt RNS szekvenálást genomra illesztők a *TopHat2* [119], a *HiSat2* [120] és a *STAR* [121] szoftverek.

Az RNS szekvenálás adatainak feldolgozása során az utolsó lépés a kifejeződési szint meghatározása. Ilyenkor általában több minta adatsorát szokták együttesen

feldolgozni. A lépés során a szoftverek először kiszámítják az egyes génekre (vagy izoformákra) illeszkedő leolvasások számát, amiket normalizálnak. Manapság a legnépszerűbb normalizálási módszerek az FPKM [122] és az RSEM számítása [123], illetve a nyers leolvasás számokból való további feldolgozása statisztikai programcsomagokkal mint például a DESeq2 és az edgeR [124, 125].

2.13.7. A DNS mutációk azonosítása

A DNS mutációk lehetnek öröklöttek, vagy újonnan szerzett (pl. tumoros) szomatikus mutációk. Azonosításuk jelentősen eltér egymástól, gyakran más a szöveti forrásuk, valamint az eloszlásuk a mintákban.

Öröklött mutációk azonosítása

Az öröklött mutációk jelen vannak minden sejtünkben, valamint tovább öröklődnek a leánysejtekben. A populáció legalább 1%-ában jelen lévő öröklött mutációkat polimorfizmusoknak nevezzük. Az öröklött mutációk lehetnek heterozigóták vagy homozigóták, attól függően, hogy a diploid sejt egyik vagy mindkét homológ kromoszómáján jelen vannak.

A mutációk azonosítása során az illesztett leolvasásokban olyan genomiális pozíciókat keresünk, ahol több leolvasásban előfordul egy változás. Mivel egy adott mutáció pozícióját általában több tíz, vagy akár több száz leolvasás is lefedi, ebből ki lehet számolni a mutáció variáns allél frekvenciáját. A számított allél frekvenciából következtethetünk arról, hogy a mutáció homozigóta vagy heterozigóta.

Napjainkban a legismertebb öröklött mutáció keresők a *samtools* [126], valamint a GATK szoftvercsomagban lévő *Unified Genotyper* és *Haplotype Caller* [111] programok. A *Haplotype Caller* előnye, hogy az illesztett adatok olvasása során a zajosnak ítélt régiókban (például ahol magas a szubsztitúciók aránya) lokális genom-összeállítás segítségével próbálja kiküszöbölni az illesztési heurisztikákból származó hibákat. A feldolgozás eredménye egy VCF (Variant Call Format, <https://samtools.github.io/hts-specs/VCFv4.2.pdf>) formátumú fájl, ami tartalmazza a mutációk pozícióját, a változások típusait, valamint a hozzájuk tartozó minőségi paramétereket.

Szomatikus mutációk azonosítása

A tumorokban lévő szomatikus mutációk azonosítása során olyan változásokat keresünk, melyek jelen vannak a tumorban, viszont a normál szövetből hiányoznak. Ehhez külön kell szekvenálni egy normál szövetből származó, illetve egy tumor szövetből származó DNS mintát. Szomatikus mutációkat azonosítani jóval bonyolultabb folyamat az öröklött mutációknál, mivel a tumorok genetikai állománya viszonylag dinamikusan változik.

A szomatikus mutációk azonosítására több szoftver érhető el, ismertebb példák a *Mutect2* [111], a *VarScan2* [127], a *SomaticSniper* [128], a *muse* [129] és a *Strelka* [130] programok. Több tanulmány bemutatta, hogy két szoftverrel azonosított mutációk 10%-80%-ban egyeznek meg [131, 132]. Az elemzés minőségének javítása érdekében a szomatikus mutációk keresését általában több szoftverrel szokták elvégezni (pl. TCGA). A kiértékelésnél végül a) csak két vagy több szoftverrel azonosított mutációkat fogadják el, vagy b) az egyes szoftverek eredményét egyesítik.

2.13.8. Kópiaszám-változás események azonosítása és a tumor arány meghatározása

Kópiaszám-változás esetén teljes kromoszómák vagy kromoszóma részek felszorzózódhatnak, vagy törölődhetnek. Ennek következménye, hogy az újgenerációs szekvenálás során a változást lefedő leolvasások száma megváltozik. Ezt a jelenséget tumoros és normál szöveti minták összehasonlításával jellemezhetik. Ilyen esetekben az elemző programok valamilyen statisztikai teszt segítségével összehasonlítják a normál és tumoros régiók közti normalizált lefedettség értékeit [133, 134].

Tumor-arány meghatározásánál azt próbáljuk megbecsülni, hogy a szekvenált DNS mekkora hányada származik tumor sejtekből. A meghatározás során általában a kópiaszám változás eseményeket, illetve a mutációs adatokat használják. A tumor arány meghatározása gyakran a normál és tumor mintában mért öröklött mutációk frekvenciájának különbsége alapján történik [135].

2.14. A túlélés-elemzés

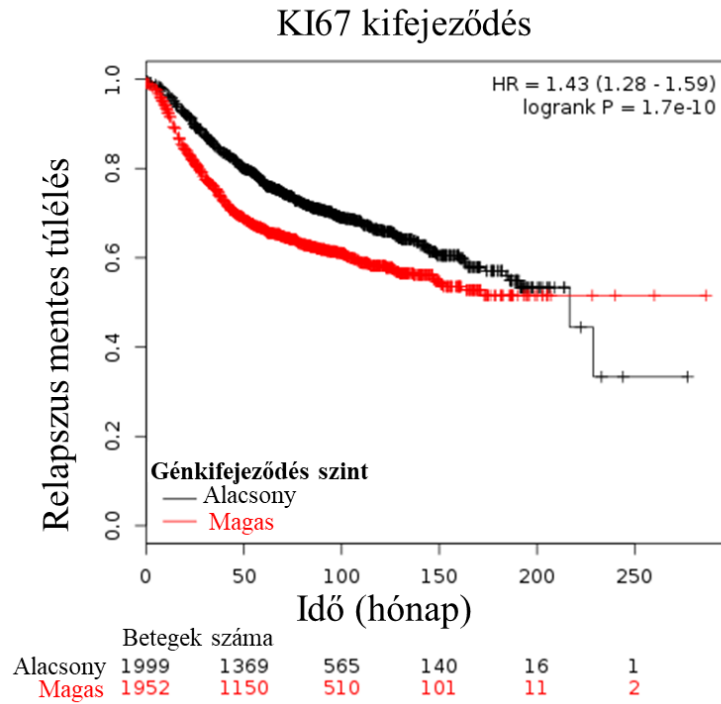
Túlélés-elemzés során valamilyen esemény bekövetkezési idejének különbségét hasonlítják össze két (vagy több) betegcsoport között. Az onkológiai kutatásban a túlélés-elemzés során leggyakrabban összehasonlított események a teljes túlélési időre, a tumor progressziómentes túlélésre, és a relapszusmentes túlélésre vonatkoznak [136]. A teljes túlélési idő a diagnózistól számított teljes élettartamot jelöli. A progressziómentes túlélés a diagnózis és a tumor progressziója közt eltelt időt jelenti. A relapszusmentes idő a diagnózis és első relapszus közt eltelt időt jelöli.

A vizsgálatokba bevont nagy betegszámok miatt gyakran nehéz a betegek pontos után követése, így az adatok hiányosak lehetnek. Erre példa, ha a beteg a vizsgálat végén még életben van, vagy esetleg a vizsgálatától függetlenül exitál. Ilyen esetekben az időt cenzoráltnak tekintik, mivel a túlélési ideje vagy ismeretlen, vagy független tényezők befolyásolták.

Kaplan–Meier görbék

A Kaplan-Meier görbék a betegek túlélését mutatják az idő függvényében [137]. A túlélési görbe y-tengelye jelöli az eseménymentes (pl. halálozott) populáció valószínűségét, az x-tengely az időt. Amíg esemény nem következik be, addig a görbe vízszintesen halad. Ahogy a betegekben bekövetkezik egy esemény, a görbe megtörik, és leesik a megfelelő százalékra. Azon esetek, ahol az idő cenzorált, azt a megfelelő időben a görbét metsző „x”-szel jelölik.

Két, vagy több beteg csoportnál az ún. logrank-próbával lehet a túlélési időket összehasonlítani [138]. A túlélés időknél eltérésének mértékét a kockázati rátával jellemzik, amit a Cox féle „arányos kockázat” modell felhasználásával szokták kiszámolni. A 12. ábrán látható a KI67 proliferációval asszociált gén kifejeződési szintjének összefüggése a betegek relapszus mentes túlélésével (*kmplot.com* [139]). Az ábrán látható, hogy a magas KI67 gén kifejeződésekor rövidebb a relapszus mentes túlélés. A két csoport közti kockázati ráta (HR) értéke 1,43; a logrank p értéke 1,7E-10.



12. ábra. A KI67 gén kifejeződési szintjének összefüggése relapszus mentes túléléssel.

3. Célkitűzések

PhD munkám során az intra-tumor heterogenitás hatását kívántam megvizsgálni *in vitro* modellezéssel, valamint petefészek-tumor mintákból származó újgenerációs szekvenálási adatsorok segítségével. Ezzel kapcsolatban olyan elemzőrendszer kidolgozását is célul tűztam ki, amely egy génmutáció hatására létrejövő génkifejeződés változásának mintázatából végezi a túlélés elemzést.

Célkitűzéseim a következők a két témával kapcsolatban:

- Megvizsgálni a tumor-heterogenitás kapcsolatát a tumorokban kimutatott szomatikus mutációkkal. Munkám során az alábbi részcélokat tűztam ki:
 - a. *In vitro* kísérletekkel vizsgálni a sejtmozgás hatását az újgenerációs szekvenálással mérhető összetételre.
 - b. Az újgenerációs szekvenálás pontosságának vizsgálata ismert összetételű sejtvonal keverékek felhasználásával.
 - c. A szekvenált tumor mérete és a detektált szomatikus mutációk száma közötti összefüggés vizsgálata petefészek-tumor minták segítségével.
- A génmutációk és a génkifejeződés közti kapcsolat vizsgálata, amihez a következő részcélokat tűztam ki:
 - a. A génmutációk eredményeképpen létrejövő génkifejeződés változás hatásának vizsgálata az emlőtumoros betegek túlélésére.
 - b. A TP53 génmutációra visszavezethető megváltozott génkifejeződési mintázat elemzése és értelmezése.

Munkám első részében célom volt tehát meghatározni az újgenerációs szekvenálás pontosságát és reprodukálhatóságát *in vitro* kísérletekkel, valamint petefészek-tumorkultúrák multirégiós szekvenálásával. Munkám második részében egy olyan elemzőrendszer kidolgozása volt a célom, amely képes a szomatikus mutációk következtében kialakuló prognosztikus génkifejeződés változásokat azonosítani, ezzel elősegítve új potenciális biomarkerek azonosítását.

4. Módszerek

4.1. *In vitro* sejtvonalas inváziós kísérletek

4.1.1. Sejtkultúrák

A sejtmotilitási mérésekhez négy melanóma sejtvonalat választottam ki: A375, MEL-JUSO, SK-MEL-28 és MEWO. A sejtvonalak közül kettő stabilan transzfektált fluoreszcens fehérjét kódoló gént tartalmazott (Linterna A375 GFP-t kifejező sejtvonal, Innoprot; MEL-JUSO FP602-GVO-CD RFP-t kifejező sejtvonal, BioCat). A sejt keveredés követése esetükben fluoreszcens videomikroszkópos kísérletekkel is el lett végezve. A MEL-JUSO sejtvonalakat RPMI-1640 médiumban, az A375 GFP, SK-MEL-28 és MEWO sejtvonalakat pedig DMEM médiumban voltak tenyésztve sejtkultúra flasksban 37°C-on és 5% CO₂ mellett. Mindkét médium ki volt pótolva 10% FBS-sel, és 1% penicillin-sztreptomycin-amfotericin B-vel.

Az inváziós kísérletek során 3 sejtvonal (A375, SK-MEL-28 és MEWO) mintáiból volt elvégezve az újgenerációs szekvenálás. Ezért ezen sejtvonalakhoz specifikus, igazolt mutációkat kerestem a COSMIC Cell Lines Project (v80) és ATCC adatbázisaiból (3. táblázat). A félkövér betűtípussal jelölt mutációkat előzetesen igazolva lettek Sanger szekvenálással.

3. táblázat. Az Ion Torrent szekvenálás során kiválasztott mutációk, amihez az AmpliSeq Designer szoftver segítségével terveztünk specifikus primereket.

Gén név	aminosav csere	cDNS	Homozigóta(2) / heterozigóta(1)	Sejtvonal
A1CF	c.913G>A	p.D305N	2	A375
ABL2	c.2795C>T	p.S932F	1	MEWO
APC	c.388A>G	p.S130G	2	SK-MEL-28
BRAF	c.1799T>A	p.V600E	2;2	A375;SK-MEL-28
CDK4	c.70C>T	p.R24C	1	SK-MEL-28
CHECK2	c.1180G>A	p.E394K	1	SK-MEL-28
EGFR	c.2257C>T	p.P753S	2	SK-MEL-28
FGFR2	c.1906G>A	p.E636K	2	A375
FLI	c.1064G>A	p.R355K	2	MEWO
JAK2	c.1774G>A	p.E592K	1	SK-MEL-28
JAK3	c.2078C>T	p.P693L	1	SK-MEL-28
MCF2	c.784G>A	p.E262K	2	A375
MYH2	c.4337C>T	p.A1446V	2	SK-MEL-28
NF1	c.4006C>T	p.Q1336*	2	MEWO
NLRP3	c.913G>A	p.D305N	2	SK-MEL-28
PAGE4	c.215G>A	p.R72Q	2	A375
PTCH1	c.2174C>T	p.P725L	2	MEWO
PTEN	c.499A>G	p.T167A	2	SK-MEL-28
RUNX1	c.908C>T	p.S303L	1	A375
SETX	c.7018C>T	p.R2340*	2	A375
TP53	c.949C>T	p.Q317*	1	MEWO
TP53	c.772G>A	p.E258K	1	MEWO
TP53	c.434T>G	p.L145R	2	SK-MEL-28
TP63	c.485C>T	p.S162L	2	MEWO
ZNF384	c.1442C>T	p.P481L	2	MEWO

4.1.1. Gyűrűs inváziós kísérletek

A gyűrűs inváziós kísérletek FlexiPERM® conB sejtkizáró szilikon gyűrűvel lettek elvégezve. A kísérletek során a gyűrűket 60x15 mm nagyságú sejtenyésző csészék

közepébe lettek helyezve. A belső sejtkultúra növesztésére alkalmas terület 3.1cm^2 , míg a külső terület 17.9cm^2 volt. A szilikon gyűrű kerülete a csésze alján volt megjelölve. Az első sejtvonal 300.000 sejtet tartalmazó 3ml DMEM médiumból volt a belső régióba helyezve. Miután a sejtek letapadtak a tenyésztő csészére, a szilikon gyűrűt ki vettük, és 1x PBS mosással eltávolítottuk a sejtörmelékeket. A második sejtvonal (minden esetben a MEWO volt) a csésze teljes felületére lett hozzáadva 6ml DMEM médiumban, amiben összesen 600.000 sejt volt. A tenyésztő táp két naponta volt cserélve. Mintavételezés során a szilikon gyűrű külső határáról gyűjtöttünk mintát a gyűrű körül egy 10mm széles steril kaparóval. Háttér mintaként a tenyésztő csésze széle körül gyűjtöttünk mintát a 10mm széles steril kaparóval.

Két fluoreszcens sejtvonal esetén videomikroszkópos felvételek is készültek. A tenyésztő csészén jelölt szilikon gyűrű határokat mikroszkóp (Leica Microsystems) segítségével folyamatosan figyeltük. Fluoreszcens mikroszkópos felvételek több pontban, a szilikon gyűrű külső és belső határában készültek. A képek feldolgozását az Image J szoftverrel végeztem.

A detektált mutáció frekvenciák összehasonlítása érdekében készítettünk kalibrációs keverékeket az A375 és a MEWO, valamint a SK-MEL-28 és a MEWO sejtvonal párokkal. A kalibrációs sorban a két sejtvonal aránya 2%, 5%, 10%, 25% és 50% voltak. Mindegyik esetben az MEWO volt a fő (major) sejtvonal.

4.1.2. DNS izolálás és minőségellenőrzés

A sejtvonal-specifikus mutációk igazolásához genomiális DNS-t izoláltunk a sejtvonal monokultúrákból. Az egyrétegű sejtkultúrát tripszin-EDTA-ás kezelés után felszuszpendáltuk, majd a genomi DNS-t a DNeasy Blood and Tissue Kittel (Qiagen, Germany) izoláltuk a gyártó protokollja alapján. A PCR reakciókhoz szükséges 500 ng DNS-t körülbelül 5×10^5 sejtől izoláltuk minden sejtvonal esetén. A szilikon gyűrű sejtinváziós esszében használt sejtekből szintén izoláltunk genomiális DNS-t. Ennek során a sejteket sejtkaparóval vettük fel. A DNS koncentrációját és tisztaságát 260 és 280 nm abszorbancia méréssel ellenőriztük Nanodrop ND1000 spektrofotométerrel.

4.1.3. Sejtvonal-specifikus mutációk igazolása

A homozigóta és a heterozigóta sejtvonal-specifikus mutációkat Sanger szekvenálás segítségével validáltuk. A sejtvonal-specifikus mutációkat a COSMIC sejtvonal adatbázisán lettek kikeresve (http://cancer.sanger.ac.uk/cell_lines). A DNS szekvenáláshoz polimeráz lánc-reakcióval amplifikáltuk a szakaszokat az izolált genomiális DNS-ből.

A PCR reakció összetételét és a PCR reakció feltételeit a 4. és az 5. táblázatba foglaltam össze. A PCR reakcióhoz a DreamTaq PCR Master Mix (K1071) lett felhasználva.

4. táblázat. A PCR reakció során alkalmazott reagensek.

Genomi DNS	500 ng
dNTP mix	10 mM
5'-3' primerek	10-10 µM
Taq polimeráz (ThermoFisher EP0701EP0701)	5 U
Taq polimeráz puffer	1x
ddH ₂ O	25 µl végtérfogatra

5. táblázat. A PCR reakció során alkalmazott hőmérséklet és idő intervallumok.

Ciklusok	Hőfok	Idő
Kezdeti denaturáció	94 °C	3 perc
35 ciklus	94 °C	3 perc
	94 °C	30 másodperc
	53 °C	30 másodperc
	72 °C	2 perc
Végső elongáció	72 °C	6 min

A PCR terméket a NucleoSpin® Gel and PCR Clean-up (740609.50) kit használatával tisztítottuk meg. A Sanger szekvenálást a Semmelweis Egyetem Genomikai

Medicina és Ritka Betegségek Intézetének a laboratóriumában végeztük. A DNS szekvenciát a BioEdit és Genedoc szoftverekkel ellenőriztük.

4.1.4. Videomikroszkópia

A sejtek mozgásának követését egy számítógép által vezérelt Leica DM IRB fordított elrendezésű mikroszkóppal végeztük Marzhauser SCAN-IM állványzattal, 10x N-PLAN objektívvel, 0,25 apertúrával, és 5,8 mm gyújtótávolsággal beállítva. Az epifluoreszcens felvételeket a mikroszkóphoz kapcsolt Zeiss Colibri világítási rendszerrel és Olympus DP70 color CCD kamerával készítettük el.

A sejtkultúrákat egy mini inkubátorban (CellMovie) tartottuk fenn a videomikroszkópia során. Az inkubátorban a sejtvonalak sejt kultúrák Petri csészékben (Greiner, Németország) 37 °C-on 5% CO₂ mellett voltak fenntartva. A rendszer 72 órán keresztül 10 percenként készített egy fáziskontraszt és egy fluoreszcens felvételt a kijelölt helyekről. A felvételeket az NIH által fejlesztett ImageJ szoftverrel dolgoztam fel.

4.1.5. Sejtmozgás mennyiségi meghatározása

A sejtmotilitás meghatározását a sejtek video-követésével végeztük. A sejteket a tenyésztés első 72 órájában követtem videomikroszkópia segítségével. Az x_i sejt $v_i(t)$ sebességét t időben az alábbi képlet (I) szerint számítottuk:

$$\text{I} \quad v_i(t) = \frac{|x_i(t+\Delta t) - x_i(t)|}{\Delta t}$$

az eltelt idő (Δt) egy órára volt beállítva. Ilyen időintervallumban az átlagos sejt elmozdulás nagyobb 10 μ m-nél, meghaladva a kézi követés során felmerülő hibákat.

Az átlagos sebességek számítását a (II) képlet szerint végeztük:

$$\text{II} \quad v(t) = \frac{1}{N(t)} \sum_{i=1}^{N(t)} v_i(t)$$

ahol $N(t)$ a követett sejtek összegét, $v_i(t)$ pedig az i -edik sejt t időbeni sebességét jelöli.

A négy sejtvonalról monokultúrában készítettünk videomikroszkópos felvételeket 72h időtartamban. Minden sejtvonalnál 20 darab sejt mozgását követtem manuálisan, majd feldolgoztam az (I) és (II) képletek segítségével.

4.1.6. Célzott szekvenálás Ion Torrent technikával

A sejtvonal-specifikus mutációkra tervezett DNS könyvtárat az AmpliSeq Designer szoftverrel terveztük összesen 25 mutációra (3. táblázat). A könyvtárat az AmpliSeq Library Kit 2.0-val készítettük el. A folyamat során a primer keveréket (pool) 10 ng genomiális DNS-hez adtuk hozzá, majd PCR reakcióval sokszorozítottuk. A szekvenálási adapterek rákötése a primerek részleges visszaemésztése után történt. A könyvtár tisztítást Agencourt AMPure XP reagenssel végeztük el, a végső koncentrációt Qubit 2.0 eszközön mértük meg. A szekvenáláshoz szükséges emulziós PCR (emPCR) reakciót és minta előkészítést Ion OneTouch eszközzel végeztük. Tisztítás után az Ion OneTouch ES készülékkel eltávolítottuk a DNS-t nem tartalmazó gyöngyöket. Szekvenálási primerek, valamint polimeráz hozzáadása után a kész Ion Sphere Particle (ISP) gyöngyöket egy Ion 314 csip-re töltöttük és Ion PGM 200 szekvenáló kittel szekvenáltuk 600x lefedettséggel.

4.1.7. Célzott újgenerációs szekvenálási adatok bioinformatikai feldolgozása

Az Ion Torrent technikával nyert célzott szekvenálási leolvasások minőségét *FastQC* szoftverrel ellenőriztem, a leolvasások nyírását a *trimmomatic* szoftverrel végeztem [112]. A megmaradt jó minőségű leolvasásokat a *BWA MEM* szoftverrel [116] illesztettem a GRCh37 referencia Humán genomra. Az illesztett leolvasásokat a *samtools* szoftvercsomaggal rendeztem és formáztam a tömörített BAM formátumra. Végül a leolvasások illesztésének minőségét tovább javítottam a GATK által javasolt eljárása alapján [111].

A mutáció keresést a *samtools mpileup* szoftver [126] alap-beállításával végeztem, ahol csak az ismert mutációkat lefedő részekre szűkítettem az elemzést. A kimenetből ki lehetett számolni az egyes mutációkat tartalmazó, vagy nem tartalmazó leolvasások arányát.

4.1.8. A szekvenálási lefedettségek összefüggésének *in silico* vizsgálata a mutációk frekvenciájával

Az *in silico* vizsgálatot elvégeztem 50x, 100x, 200x, 250x, 400x, 500x, 600x, 700x és 1000x lefedettségi értékekkel, valamint minden percentilisen 1-99% között. A szimulálás során először generáltam 10.000 elemet (leolvasást). Az adott percentilisnek

(várt) megfelelő arányú elemet megváltoztattam, majd az elemeket összekevertem Fisher-Yates keveréssel. Végül az első N elemből kiszámítottam a mutáció frekvenciát (mért), ahol N az adott lefedettséget képviselte. Minden lefedettségi értékre 100x ismétlést végeztem.

4.2. Az ovárium-tumor minták újgenerációs szekvenálási adatainak feldolgozása

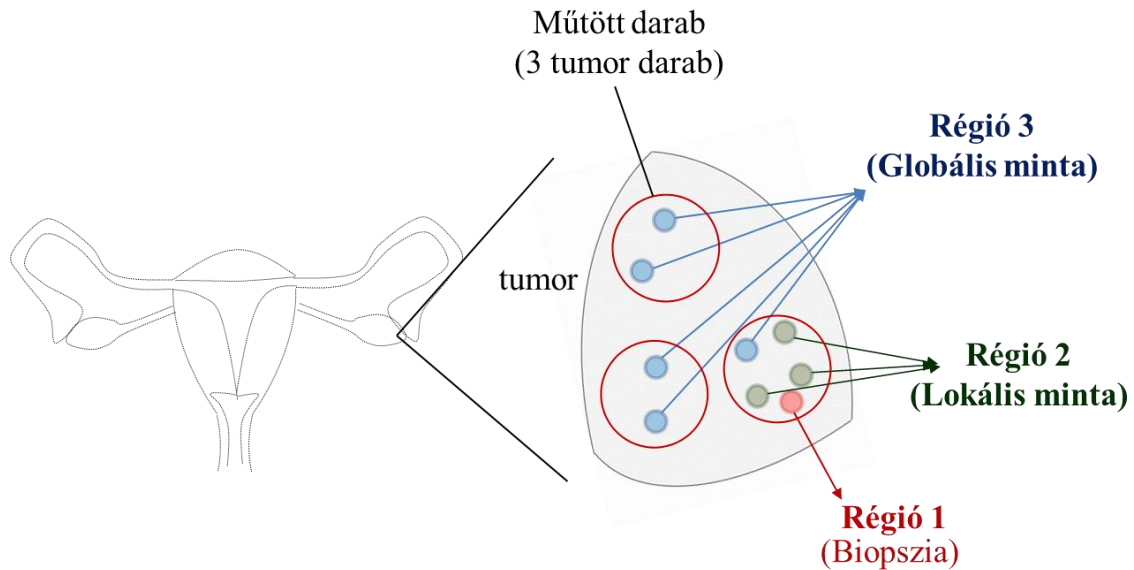
A petefészek tumorok multi-régió szekvenálása öt betegből volt elvégezve. A minták gyűjtése az Országos Onkológiai Intézetben történt a 001-ÁLT.-9444-1/2013/59 kódú etikai engedéllyel. Minden betegtől egy normál vérszövet, valamint a műtött tumorból három részlet lett kigyűjtve. Az egyes mintákra vonatkozó klinikai adatokat a 6. táblázatba foglaltam össze. Fontos volt, hogy egyik betegnél sem alkalmaztak genetikai összetételt befolyásoló neoadjuváns kezelést

6. táblázat. A petefészek tumoros betegek fő klinikai jellemzői.

Azonosító	Kor	Stádium	Grade	Hisztológia	CA125 (U/mL)	Laparotómia (optimális/szub-optimális)
909	64	pT3a	G3	szerozus	326	Optimális
847	65	pT1c	G1	endometrioid	295	Optimális
328	64	pT3c	G3	szerozus	459	Szub-optimális
358	60	pT3b	G3	szerozus	1970	Optimális
239	49	pT3c	G1	endometrioid	672	Szub-optimális

A vér és tumor mintákból a „Qiagen DNeasy Blood and Tissue Kit” kit segítségével izoláltuk a DNS-t a gyártó protokollja alapján (Qiagen GmbH, Hilden, Németország). A DNS koncentrációk NanoDrop ND-1000 spektrofotométerrel (BCM, Houston, TX, USA) és Qubit 3.0 Fluorometerrel (Thermo Fisher Scientific, USA) is lemértük.

Minden beteg tumor mintájából három térben elválasztott régióból izoláltunk DNS-t, amit a 13. ábrán illusztrálok. A „Régió 1” minta egy biopszia mintának felel meg. A „Régió 2” a lokális minta, ekkor a biopszia közvetlen közelében lévő 3 darabot egyesítettünk, amiből elvégeztük a DNS izolálást. A „Régió 3” a globális mintának felel meg, ekkor a tumornak mind a három részéből együttevve izoláltunk DNS-t. A betegektől rendelkezésre állt vérminta is az öröklött mutációk azonosítására.



13. ábra. Mintavételezés a tumor mintákból. Minden betegből 3 tumor részlet állt rendelkezésre. A biopsziánál és lokális mintáknál a DNS-t egy tumor részletből izoláltuk. A globális minta esetén mind a három részletből lett DNS izolálva.

Az egyes mintákból izolált DNS koncentrációkat és mennyiségeket a 7. táblázat foglalja össze. A minta azonosító első három karaktere a beteg azonosító, utolsó karaktere jelöli, hogy a vérmintából (0), a biopsziából (1), a lokális mintából (2), vagy a globális mintából (3) származik a DNS. Ezeket a mintákat a BGI laboratóriumába (<https://www.bgi.com/>, Hong Kong, Kína) küldtük ki teljes exom szekvenálásra.

7. táblázat. Újgenerációs szekvenálásra kiküldött minták koncentrációja

Minta azonosító	koncentráció (Qubit) (ug/ml)	térfogat (ul)	DNS hozam (ug)	OD 260 /280
909-0	58,45	100	5,845	1,85
909-1	75,6	200	15,12	1,92
909-2	70,6	200	14,12	1,98
909-3	120	200	24	1,96
847-0	125,6	100	12,56	1,81
847-1	100	100	10	1,98
847-2	93,2	100	9,32	1,98
847-3	136	>50	6,8	1,98
328-0	66,35	100	6,635	1,83
328-1	177,6	100	17,76	1,97
328-2	99,8	100	9,98	2,04
328-3	99,4	100	9,94	2,04
358-0	25,2	100	2,52	1,53
358-1	507,8	100	50,78	1,91
358-2	133,6	100	13,36	1,93
358-3	123,6	100	12,36	1,92
239-0	45,1	100	4,51	1,8
239-1	122,8	100	12,28	1,99
239-2	72	100	7,2	1,98
239-3	188	100	18,8	1,98

4.2.1. Leolvasások illesztése a referencia-genomra

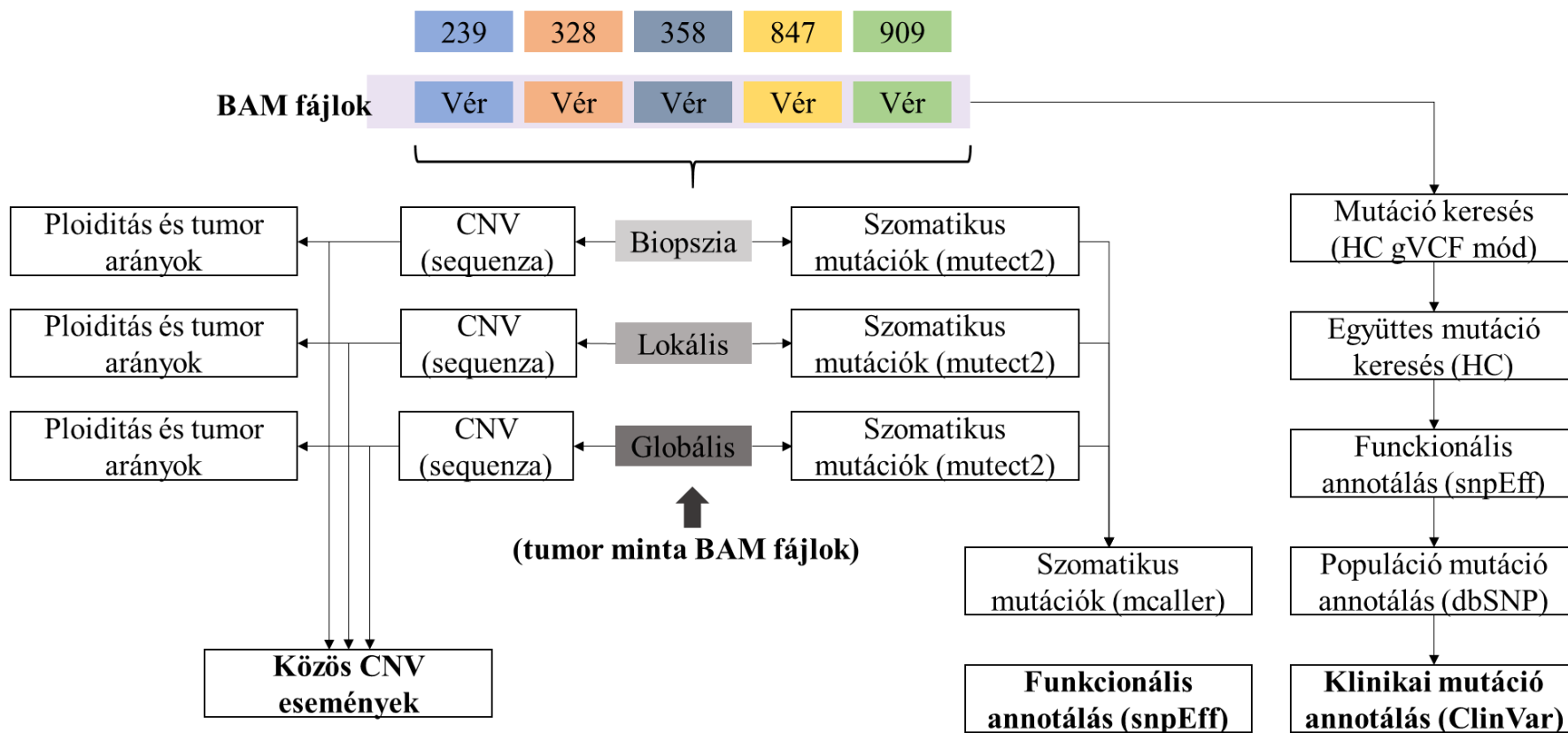
Az exom szekvenálásból származó leolvasások minőségét a *FastQC* szoftverrel ellenőriztem, a leolvasások nyírását a *trimmomatic* szoftverrel végeztem. A megfelelő minőségű leolvasásokat a *BWA MEM* szoftverrel illesztettem a TCGA lerakatából letöltött GRCh38.d1.vd1 Humán referencia genomra. A SAM formátumú illesztett leolvasásokat a *samtools* szoftvercsomaggal rendeztem és formáztam a tömörített BAM formátumra.

4.2.2. Az illesztések előfeldolgozása

Az illesztett és rendezett BAM-formátumú illesztett leolvasásokban a duplikátumokat a *picard-tools* szoftver segítségével jelöltem meg. Az illesztett leolvasásokban az inszerciók és deléciók környékén újraillesztést végeztem a GATK *RealignerTargetCreator* és *IndelRealigner* programok segítségével. Ez az újraillesztés kijavítja az illesztés minőségét azon régiókban, ahol magas az indelek és szubsztitúciók aránya. Végül a leolvasásokon a minőségi értékeket újra kalibráltam a GATK *BQSR* szoftverével.

4.2.3. A szomatikus és az öröklött mutációk azonosítása és annotálása

A szomatikus mutációk azonosítását a GATK *mutect2* szoftverrel végeztem el (14. ábra). A *mutect2* egyszerre elemez egy normál szövetből és egy tumor szövetből származó BAM fájlokat. Működése során kikeresi azon mutációkat, melyek azonosíthatók a tumor mintában, viszont a normál mintában hiányoznak. A mutáció keresést alapbeállításokkal végeztem. A szoftver eredményét csak azokra a *mutect2* által elfogadott („PASS”-ként jelölt) mutációkra szűrtem, ahol mind a normál, mind a tumor mintában is legalább 5 leolvasás lefedett, a mutációt a tumorban legalább 5 leolvasás tartalmazta, illetve a normál mintában maximum 1 leolvasásban volt jelen.



14. ábra. A petefészek tumorok adatainak feldolgozása. A tumor ploiditás és tumor sejt arány számítást a sequenza szoftverrel, szomatikus mutációk azonosítását a mutect2 és az mcaller szoftverekkel, az öröklött mutációkat pedig a GATK HaplotypeCaller szoftverrel azonosítottam.

Az öröklött mutáció azonosítását a GATK *HaplotypeCaller* programjának alap beállításával végeztem (14.ábra). Az egyes normál minták illesztett és rendezett BAM fájljaiból gVCF módban történt az elsődleges mutáció keresés. A betegekre külön generált GVCF fájljokból együttesen genotipizáltam a GATK *GenotypeGVCFs* program segítségével. A nyers mutációkat tartalmazó VCF fájlban végül a *VariantRecalibrator* programmal végeztem a minőségi kalibrálást [111].

A mutációk génekhez való kapcsolását az *SNPeff* programmal [140] végeztem a GRCh38.86 Humán genom verzió adatbázisának felhasználásával. Ez alatt csak a kanonikus gén izoformákat használtam az annotálásra. A zaj arányának csökkentése érdekében csak génekben, vagy valamely géntől maximum 5,000 bp távolságra elhelyezkedő mutációkat fogadtam el.

Végül a normál minták mutációit tovább annotáltam a *dbSNP* [141] és *ClinVar* [142] adatbázisok felhasználásával. A *dbSNP* adatbázissal való annotálás segítségével sikerült beazonosítani az irodalomban már ismert populációs polimorfizmusokat. A *ClinVar* adatbázis felhasználásával ki tudtam keresni az ismert patogén, illetve benignus mutációkat.

4.2.4. A szomatikus mutáció együttes azonosítása

A *mutect2* csak egyedi minták feldolgozására alkalmas, emiatt kifejlesztettem egy mutáció kereső szoftvercsomagot a minták együttes feldolgozására, amit *mcaller*-nek neveztem el (14.ábra). A szoftver a megadott bemeneti mutáció listán egyenként halad végig, a BAM fileokban kikeresi azon „aktív” régiókat, ahol valamennyi mintában 1-nél több mutáció található, amiknél elvégzi a mutáció keresést. Az eredmény egy olyan táblázat, ami tartalmazza a szomatikus mutációkat, valamint minden mintára egyedileg listázza a minőségi paramétereket, mint a lefedettség, a mutáns leolvasások számát, a szignifikancia szintet a háttérhez képest, és egyéb minőségi paraméterek. A program a *github* lerakatban érhető el (<https://github.com/pongorlorinc/mcaller>).

4.2.5. A leolvasások szűrése és beolvasása az *mcaller* szoftverben

Az *mcaller* a feldolgozást a rendezett illesztési adatokat tartalmazó BAM fájljokból végzi. Az egyes leolvasások szűrési paraméterei a szoftverben külön változtathatók.

Alapértelmezett használatokor kiszűri azon leolvasásokat, amikben 1) egy vagy kisebb az illesztési minőség, 2) duplikátumként van jelölve, 3) paired-end szekvenálásnál a pár másik kromoszómára illeszkedik, 4) négy vagy több változás van a referenciához képest, valamint 5) találhatóak ún. „hard” valamint „soft” levágások az illesztésben.

Az elfogadott leolvasások feldolgozása során az egyes bázisok adatait a program kétdimenziós minőségi mátrixokban tárolja, külön minden bementi mintára. Egy mátrix x-tengelye a vizsgált régió referencia szerinti pozícióját jelöli, y-tengelye viszont az adott pozícióban illeszkedő bázisok minőségét tartalmazza a négy lehetséges bázisra vetítve. Mivel megkülönböztetjük az előre- valamint vissza irányban illeszkedő leolvasásokat, emiatt a két irány adatait külön mátrixban tárolja a program. A beolvasás során az *mcaller* összehasonlítja az összes leolvasás szekvenciáját a referenciával és egy külön vektorban eltárolja minden pozíciónál a nem-referencia változások számát. Ha egy adott pozícióban elegendő változás található, azt eltárolja egy szótárban (angolul „hash”), mint mutáció keresésre alkalmas aktív régió.

4.2.6. Mutációk keresés aktív régiókban

A mutáció keresés során a szoftver először az aktív régió +/-10bp környezetében kiszámolja a zaj arányát a pozíciónként található mutációk átlaga alapján. A zajhoz továbbá hozzáadja a lefedettségének 0,25%-át, ami közelíti a szekvenálás várható pontosságát. A szoftver *Poisson* kumulatív eloszlás függvény felhasználásával számítja az aktív régióban lévő mutáció valószínűségét a zajhoz képest.

A következő lépésben a szoftver megvizsgálja a mutáció irányultsági eltérését az adott régiót lefedő leolvasásokhoz képest. A lépés során a szoftver kiszámítja az adott régió leolvasási irányultságait, valamint a mutáns leolvasások irányultságait a minőségi mátrixból. A kapott értékeket egy *Fisher* teszt segítségével hasonlítja össze.

Az utolsó lépés során a szoftver megvizsgálja, hogy a mutáns leolvasások minősége mennyire tér el a környezetben lévő leolvasások minőségétől *Poisson* kumulatív eloszlás függvény felhasználásával. Ez segít kiszűrni a rossz minőségű leolvasásokból származó mutációkat.

A mutáció keresés végén az *mcaller* minden mintára külön oszlopban kilistázza minden mutációra az egyes tesztek szignifikancia értékét. A szoftver normál mintát is

feldolgoz a tumor mintákkal együtt, így eredményében feltünteti az öröklött mutációkat is.

4.2.7. Kópiaszám változás elemzés

A kópiaszám változás elemzést a *sequenza* szoftver segítségével végeztem (14. ábra). A *sequenza* részben python, részben R szoftvernyelven készült, előfeldolgozásként a *samtools* és *copynumber* (R csomag) programokat használja fel. Bementeként két BAM fájlt kell megadni, egyet-egyét a normál és tumor mintákból.

Az elemzés első lépéseként a BAM formátumú leolvasásokat *mpileup* formátumra kell alakítani a *samtools* szoftvercsomaggal. Ez alatt kiszűrtem az 5-nél kisebb illesztési minőséggel rendelkező leolvasásokat. Az *mpileup* kimenetből a szoftver először kikeresi a normál mintában lévő mutációkat (heterozigóta és homozigóta), valamint kiszámítja a variánsok allél frekvenciájukat. A következő lépésben kiszámítja ezen mutációk variáns allél frekvenciáját a tumorban is. Ez a lépés segít beazonosítani azon genom régiókat, ahol a mutációk frekvenciája eltolódik a két minta között egy feltehető kópiaszám változás esemény miatt. Következő lépésben GC-arány normalizálást végez a normál és tumor minták lefedettségi arányaival. A normalizálás után a *copynumber* csomaggal elvégzi az allél-specifikus szegmentálást, valamint a tumor arány és a ploiditás becsléseket.

A szoftver által választott optimális értékek gyakran alul becsülik a tumor arányt, és felülbecsüli a ploiditást. Az eredmények javítása érdekében minden elemzést külön elvégeztem a szuboptimális értékekkel is. A ploiditást és a tumor arányokat az egy beteg mintái közt legközelebbi értékek alapján választottam.

4.2.8. A mutációs mintázatok meghatározása

A mutációs mintázatokat a *MutationalPatterns* programcsomaggal [143] számítottam ki R környezetben. Először minden beteg adataiból kiválasztottam az egyes régiókban közös mutációkat. A következő lépésben letöltöttem a BSgenome által előkészített GRCh38 referencia humán genomot. Ez után letöltöttem az Alexadrov és munkatársai által előre generált mintázatokat [26] a COSMIC adatbázisból [17]. A letöltött adatok felhasználásával a *MutationalPatterns* csomagban lévő

fit_to_signatures() és *plot_contribution()* függvényekkel határoztam meg az egyes mintázatok hozzájárulását a betegekben mért mutációkhoz.

4.3. A TCGA adatbázis emlődaganatos és nem-kissejtes tüdőrák mintáinak feldolgozása

A TCGA (The Cancer Genome Atlas) adatbázisból összesen 762 emlőtumoros és 555 nem-kissejtes tüdőrákos beteg adatait dolgoztam fel. Az adatokat a *geneTorrent* (v3.8.5) szoftver segítségével töltöttem le a CGhub lerakatról. Minden betegnél adva volt egy tumorszöveti, valamint egy normál vérből végzett exom szekvenálási adatsor. A feldolgozott adatbázis meghaladta a 20Tb méretet.

Mivel a letöltött exom szekvenálási adatok a GRCh37, GRCh37-lite, és HG19 genom verziókra lettek illesztve, emiatt az adatelemzést csak a fő kromoszómákra szűkítettem, mivel a bázissorozatuk minden esetben megegyezett, csak a kromoszóma nevek különböztek (példa. „chr1” és „1”).

4.3.1. Azonosított mutációk annotálása

Az exom szekvenálási adatokból való mutációk azonosítását a *Mutect* szoftver alap beállításával [144] végeztem. A *Mutect* osztályozó rendszere által „KEEP”-ként jelölt mutációkat használtam fel. További szűrés során kikerestem azon mutációkat, ahol a lefedettség legalább 20x, valamint legalább 4 leolvasás támasztotta alá a mutációt.

A megszürt mutációk annotálását az *SNPeff* (v3.5) szoftverrel végeztem [140]. Ehhez a GRCh37 Humán genom annotációs adatbázisát az *SNPeff* letöltő moduljával készítettem elő. Az *SNPeff*-el végzett annotáció eredményeként minden mutációról megkaptam, hogy mely génben, vagy gén közelében helyezkedik el, illetve, hogy milyen típusú változást okozott. A kapott változás típusok az alábbi kategóriákba voltak sorolva: aminosav cserét okozó, csendes mutáció, STOP vagy START kodont érintő mutáció, intron mutáció, splice mutáció, valamint korai STOP kodont okozó mutáció.

Az annotált *SNPeff* mutációkat egy betűs kódolás segítségével redukáltam. Az elemzések egyszerűsítése érdekében készítettem egy olyan adat mátrixot, amiben a betegek oszlopokban, a gének sorokban lettek feltüntetve, a mutáció típusok a cellákban az egy betűs kódolással szerepeltek.

4.3.2. Génkifejeződési adatok gyűjtése és normalizálása

A TCGA adatbázisban minden betegről elérhető volt a tumor minta RNS-szekvenálásból származó normalizálatlan génkifejeződési adata is. Az RNS szekvenálást minden esetben Illumina HiSeq 2000 eszközön végezték. Az adatok normalizálását RSEM-el végeztem [123].

A microarray alapú génkifejeződési adatokat az EGA (European Genome-Phenome Archive) és GEO (Gene Expression Omnibus) adatbázisokból értem el. Az adatbázisban 5,934 beteg nyers Affymetrix CEL formátumú génkifejeződési fájlja volt elérhető. Az adatokat MAS5 normalizálással dolgoztam fel a *Bioconductor Affy* [145] könyvtárában R (v3.02) szoftverkörnyezetben.

4.3.3. Statisztikai elemzés

A génkifejeződés változást ROC (Receiver Operating Characteristic) elemzéssel végeztem R környezetben a *Bioconductor ROCR* csomagjával [146]. Az AUC (Area Under the Curve) értéket a *ROCR prediction()* és *performace()* függvények felhasználásával számítottam ki. A számított eredményeket $p < 0,01$ valamint $> 0,65$ AUC értékekkel szűrtem.

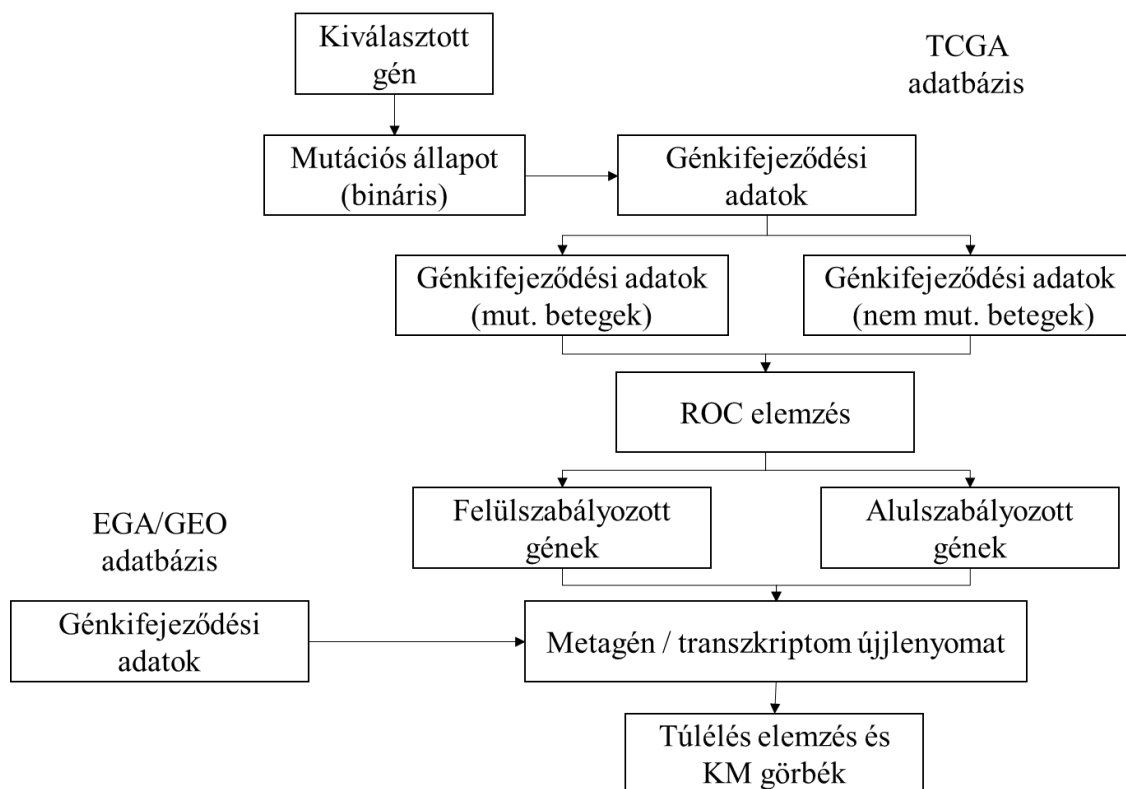
Túlélés elemzést az R környezetben fejlesztett *survplot* csomaggal végeztem (<http://www.cbs.dtu.dk/~eklund/survplot/>). Szignifikancia küszöbként a $p < 0,05$ valamint $HR > 1,44$ választottam.

4.3.4. A G-2-O (Genotype-to-Outcome) elemző rendszer

Munkám során egy olyan elemző rendszert fejlesztettem, amely párhuzamosan gén kifejeződési és génmutációs adatokat is felhasznál a túlélés elemzés során. Az algoritmus három adatbázist használ fel: a TCGA adatbázisból származó mutációs és RNA-seq adatokat, valamint független validációs adatbázisként egy microarray alapú génkifejeződési adatbázist. A módszer működése két fő szakaszra bontható, bemenetként egy gén nevet vagy azonosítót kell adni, kimenetként meg Kaplan-Meier túlélési görbét készít (15. ábra).

Az elemzés első szakaszában a szoftver kikeresi a kiválasztott gén mutációjának állapota függvényében megváltozott kifejeződést mutató géneket a TCGA adatok felhasználásával. A mutációval rendelkező és vad típusú tumorok közti génkifejeződés változást ROC (Receiver Operating Characteristic) elemzéssel végzi el a szoftver, aminek az eredménye egy táblázat, ami tartalmazza a felül- és alul szabályozott géneket, a két csoport medián génkifejeződési szintjét, valamint a statisztikai eredményeket.

A második szakaszban történik a túlélés elemzés a független microarray adatbázis felhasználásával. Az előző lépésben azonosított géneket a szoftver kikeresi a microarray adatbázisban, majd kiszámolja minden betegnél a gének átlagos kifejeződési szintjét (ujjlenyomat kifejeződés). A kapott metagén kifejeződési szintekből kikeresi a medián értéket, amely alapján a betegeket magas vagy alacsony kifejeződési csoportokra bontja. Végül a kettébontott csoportok felhasználásával készít túlélés elemzéssel Kaplan-Meier ábrákat. Az elemzés során lehetőség van kiválasztani, hogy a metagén legyen kombinált (minden gén), vagy ketté bontva alul- illetve felül szabályozott génekre [147]. A metagén ketté bontásakor két túlélési görbe készül, külön az alul- illetve a felül szabályozott génekre.



15. ábra. A G2O elemzőrendszer folyamatábrája. A rendszer először azonosítja a kiválasztott gén mutációjával összefüggő megváltozott kifejeződést mutató géneket ROC elemzéssel a TCGA adatbázis felhasználásával, majd az azonosított génekkel túlélés elemzést végez egy független gén csip adatbázison.

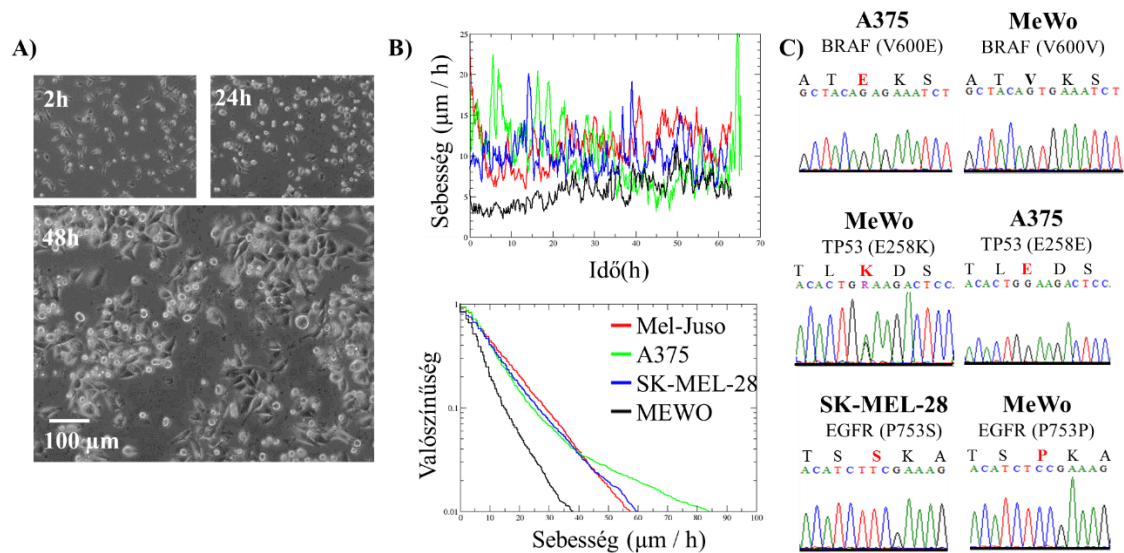
5. Eredmények

5.1. A sejtvonalak motilitásának hatása az újgenerációs szekvenálás eredményére

5.1.1. Sejtvonal paraméterek meghatározása monokultúrákkal

A négy vizsgált sejtvonalról (A375, MEL-JUSO, SK-MEL-28 és MEWO) készítettünk egy-egy monokultúras videomikroszkóp-felvétel sorozatot. Ennek során 72 órán keresztül 10 percenként készültek a fáziskontraszt felvételek (16/A ábra). Az elkészült felvételekből minden sejtvonal esetén 20 darab sejtnek a mozgását követtem, majd elemeztem. Az elemzés eredményeként megkaptam a négy sejtvonal mozgási sebességét, valamint a sebesség valószínűségi eloszlásokat (16/B ábra). Az eredményekből látható, hogy a MEWO sejtvonal általában 2-5 $\mu\text{m}/\text{óra}$ -al lassabb volt a többi sejtvonalnál, így ezt választottam kontroll sejtvonalnak. Inváziós sejtvonal kategóriába az A375 és SK-MEL-28 sejtvonalakat soroltam be. A MEL-JUSO (GFP) sejtvonalat kontrollként használtam fel a fluoreszcens videomikroszkópos követési kísérleteknél.

Az újgenerációs szekvenáláshoz kiválasztott sejtvonalakból (A375, SK-MEL28, MEWO) Sanger szekvenálással leellenőriztünk 3 sejtvonal specifikus mutációt (16/C ábra). A BRAF gén V600E mutációja jelen volt mind az A375, mint a MEWO sejtvonalakban, az SK-MEL-28 sejtvonal viszont nem tartalmazta. A TP53 gén tekintetében az SK-MEL-28 sejtvonal szintén vad típusú volt, miközben az A375 sejtvonal tartalmazott egy szinonim mutációt, illetve a MEWO sejtvonalban egy aminosav cserét okozó mutáció volt detektálható. Ezen kívül az EGFR génben kimutattunk két különböző aminosav cserét okozó DNS mutációt az SK-MEL-28 és MEWO sejtvonalakban.



16. ábra. Az inváziós kísérletek előtt meghatározott sejtvonal specifikus paraméterek. A) A monokultúrás videomikroszkóp felvételek az A375 sejtvonalról. B) A videomikroszkóp felvételekből meghatározott mozgási sebességek és sebességi valószínűségek. C) Sanger szekvenálással igazolt mutációk a BRAF, a TP53 és az EGFR génekben.

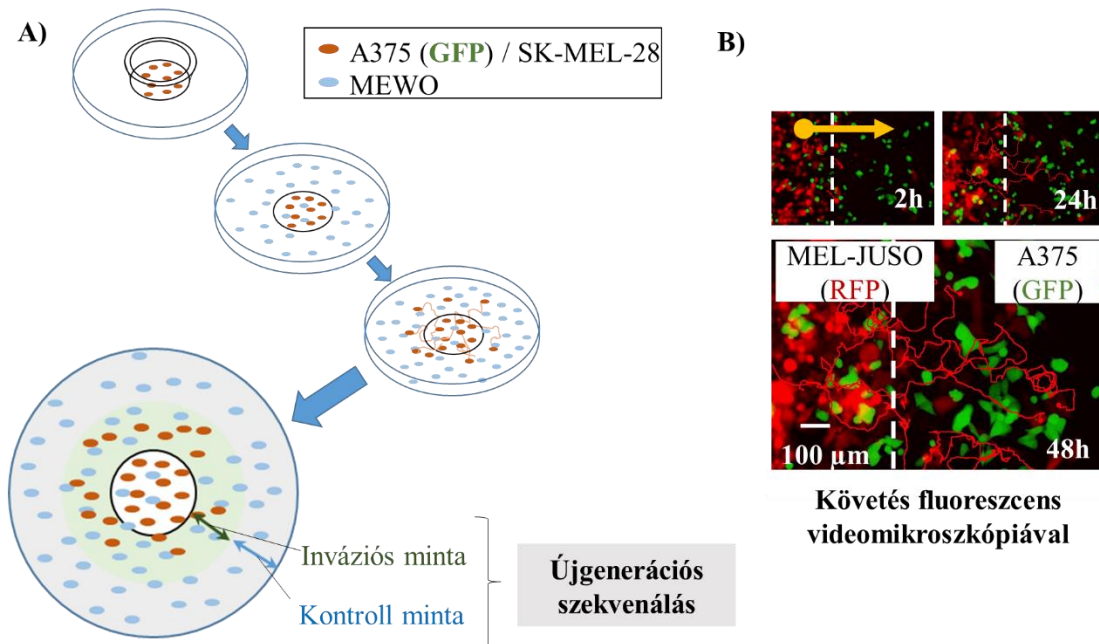
5.1.2. Az inváziós kísérletek

Az inváziós kísérleteket szilikon gyűrűk felhasználásával végeztük (17/A ábra). A kísérlet előkészítése során először az inváziós sejtvonalat (A375 vagy SK-MEL-28) tenyésztettük a szilikon gyűrűn belül. Miután a sejtek megfelelően letapadtak, eltávolítottuk a gyűrűt, és a csésze teljes felületére helyeztük le a MEWO sejtvonalat.

Az invázió vizsgálatát ellenőrzésként fluoreszcens videomikroszkópia segítségével is elvégeztük. Ennek során a MEL-JUSO (RFP) sejtvonalat választottuk, mint inváziós sejtvonal, az A375 (GFP) sejtvonal volt a kontroll. A felvételekben 20 sejt követésével vizsgáltam a vándorlási irányokat. Az útvonalakból az látszódott, hogy bár többnyire véletlenszerű volt a mozgás irány, valamennyire a tér telítettségétől függött, ahol a sejtek inkább a kevésbé telített részek irányába mozogtak.

Az inváziós kísérletekhez az A375 / MEWO és SK-MEL-28 / MEWO sejtvonal párokat használtuk, mindkét esetben a MEWO volt a külső sejtvonal. Mindkét kísérletet kétszeres ismétléssel végeztük el. A 72 órás inváziós kísérlet után sejtkaparóval 10mm széles sávban gyűjtöttünk mintát a szilikon gyűrű széle körül és kontrollként a tenyésztő

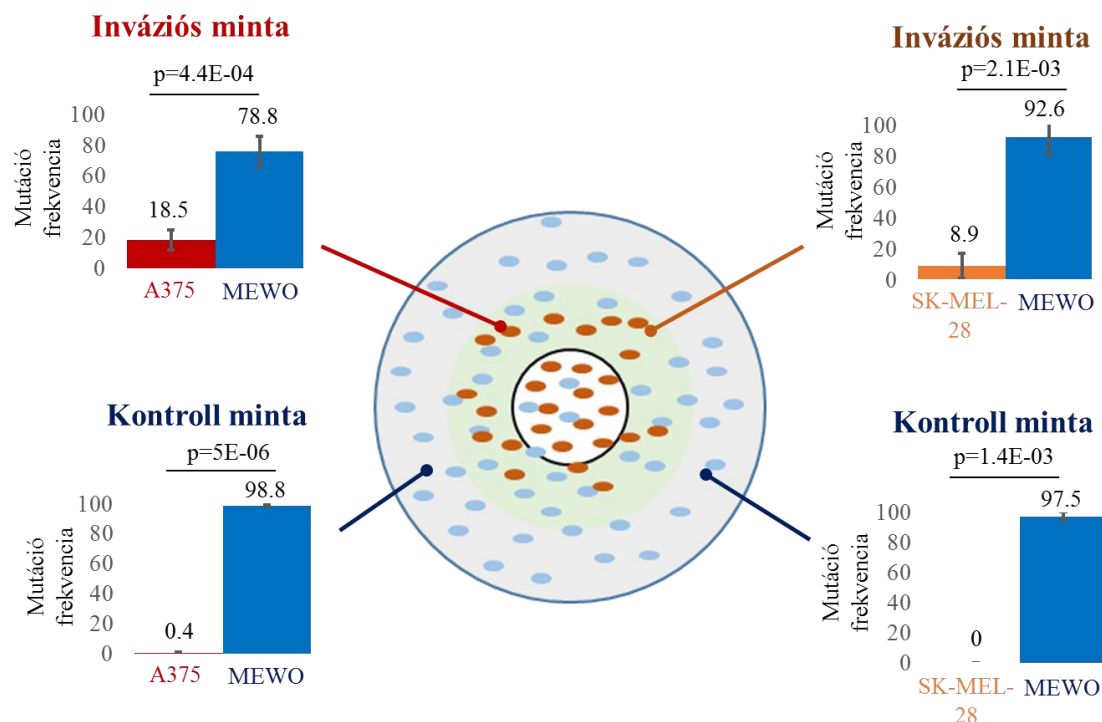
csésze fala körül. A mintákból végzett DNS izolálást követően végeztük a 25 gén szakasz célzott újgenerációs szekvenálását (3. táblázat). Az azonosított DNS mutációk típusa és frekvenciája alapján mértem az invázió mértékét.



17. ábra. A gyűrűs inváziós kísérlet elve. A) Először az egyik sejtvonalat tapasztjuk le egy szilikon gyűrű segítségével a tenyésztő csésze közepére. A letapadás után a gyűrűt eltávolítjuk, és a második sejtvonalat a teljes felületre tapasztjuk. Az inváziós kísérlet végén két mintát gyűjtünk a szilikon gyűrű körül. **B)** Az inváziós kísérletek fluoreszcens videomikroszkópiával való validálása az A375 (RFP) és a MEL-JUSO (GFP) sejtvonalakkal.

Az A375 és a MEWO sejtvonalak inváziós kísérlete során az inváziós mintákban 18,5% volt az infiltráció mértéke (18. ábra). Az értéket a két sejtvonal homozigóta mutációk frekvenciája alapján számítottam ki. Ehhez képest a külső kontroll régióban kevesebb, mint 1% volt az infiltráció mértéke, ami alacsony zaj arányra utal. A MEWO sejtvonal heterozigóta mutációinak frekvenciája 54% volt a kontroll mintában, valamint 42% az inváziós mintában.

Az SK-MEL-28 és a MEWO sejtvonalak inváziós kísérlete során 8,9% volt az infiltrációja mértéke az inváziós mintákban (18. ábra). A kontroll régióban 2 leolvasásban azonosítottam SK-MEL-28 sejtvonal specifikus mutációkat, ami szinte teljesen elhanyagolható volt a közel 1000 db leolvasáshoz képest.



18. ábra. Az inváziós kísérletek eredményei. Az A375 / MEWO sejtvonalak inváziós kísérlete során 18,4% volt az invázió mértéke a belső mintában, külső mintában 0,4% kontaminációt mértünk. Az SK-MEL-28 / MEWO sejtvonalak inváziós kísérlete során 8,9% inváziót mértem a belső mintában, a külső kontroll mintában 0% volt a kontamináció.

A külső kontroll régióban lévő MEWO sejtvonal homozigóta mutációk frekvenciája egyik esetben sem érték el a 100%-os értéket. Ráadásul az inváziós sejtvonalak homozigóta mutációk frekvenciájának a szórásai egy kísérleten belül 6-12% között ingadoztak. Ez erősen befolyásolhatja a mutáció detektálásának lehetőségét, valamint a mutáció frekvenciákra épülő tumor evolúciót becselő programok eredményét. A biológiai ismétlések között átlagosan 3,8% (0-9% intervallumban) volt a mutáció frekvenciák közti különbség magas lefedettségű mutációknál (>100x leolvasás), átlagosan 7,6% (0-34% intervallumban) ha az alacsonyabb lefedettségű mutációkat is bevontam.

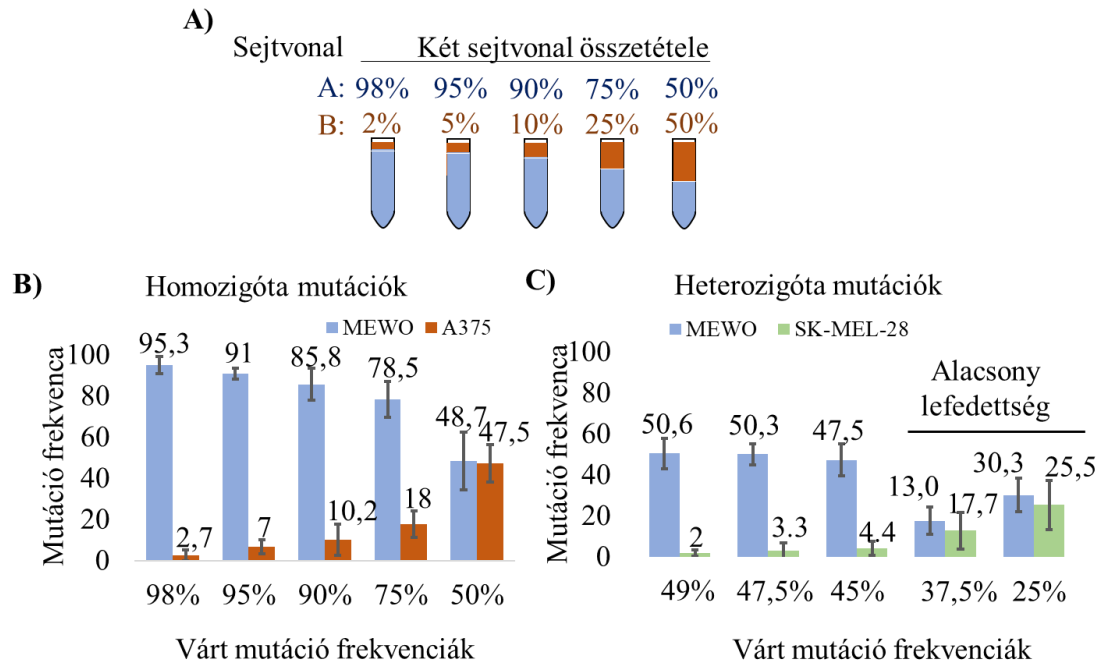
Invázió során az A375 sejtvonal szignifikánsan nagyobb infiltrációt ért el az SK-MEL-28 sejtvonalhoz képest ($p=0,011$), ami a MEWO sejtvonal homozigóta mutáció frekvenciáinak összehasonlítása során is észlelhető ($p=0,0059$) volt. Viszont nem tapasztaltam szignifikáns különbséget a két sejtvonal inváziója között, amikor csak heterozigóta mutációkat hasonlítottam össze ($p=0,39$). Ennek egy lehetséges magyarázat,

hogy a heterozigóta mutációk kisebb mutáció frekvenciával rendelkeznek, ezzel csökkentve az százalékos különbségeket.

5.1.3. Sejtvonal kalibrációs sorok elemzése

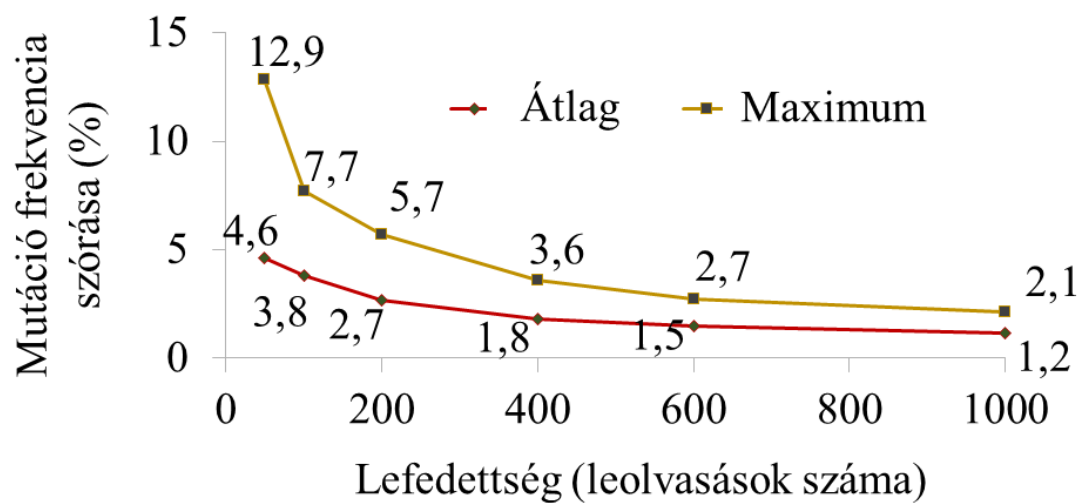
Kalibrációs sort vettünk fel annak vizsgálatára, hogy ismert összetételű mintában milyen pontosan lehet meghatározni a sejtvonalak arányát újgenerációs szekvenálással. A kalibrációs szekvenálás során két sejtvonal lett összekeverve 2% / 98%, 5% / 95%, 10% / 90%, 23% / 75% és 50% / 50% arányokban (19/A ábra), az A375 / MEWO és SK-MEL-28 / MEWO sejtvonal párokkal külön, mindkét esetben két technikai ismétléssel.

Azon esetekben, ahol az egyik sejtvonal tette ki az arány nagy részét, ott a mutáció frekvenciák nagyobb pontossággal követték az összetétel arányát. Viszont, ahol a két sejtvonal (közel) azonos mennyiségben volt jelen, ott a mutációk frekvenciájának a szórásai 8% és 17% közé estek magas lefedettségeknél (19/B ábra). Heterozigóta mutációknál szinte minden esetben magasabb volt a szórás, ilyenkor a két sejtvonal aránya kevésbé befolyásolta az eredményt. A mutációk frekvenciájának a szórásai 1,6% és 5,6% közé estek a technikai ismétlések között is (19/C ábra).



19. ábra. A sejtvonal kalibrációs sor kísérletek során újgenerációs szekvenálással mért mutáció frekvenciák. A) A kalibrációs sor összetétele. B) A MEWO / A375 sejtvonal párokból mért homozigóta mutáció frekvenciák. C) A MEWO / SK-MEL-28 sejtvonal párokból mért heterozigóta mutáció frekvenciák.

Az *in silico* modellezés során számított mutációk frekvenciájának átlagos szórása a kalibrációs szekvenáláshoz hasonló eredményt mutatott (20. ábra). Alacsony lefedettségeknél (50x) az átlagos szórások 4.6% voltak, míg magas lefedettségeknél ezek 1,2%-ra csökkentek.



20. ábra. A mutáció frekvencia és szekvenálási lefedettség összefüggésének *in silico* vizsgálata. A lefedettség növelésével csökkentek a maximális, valamint az átlagos mutáció frekvenciák szórása.

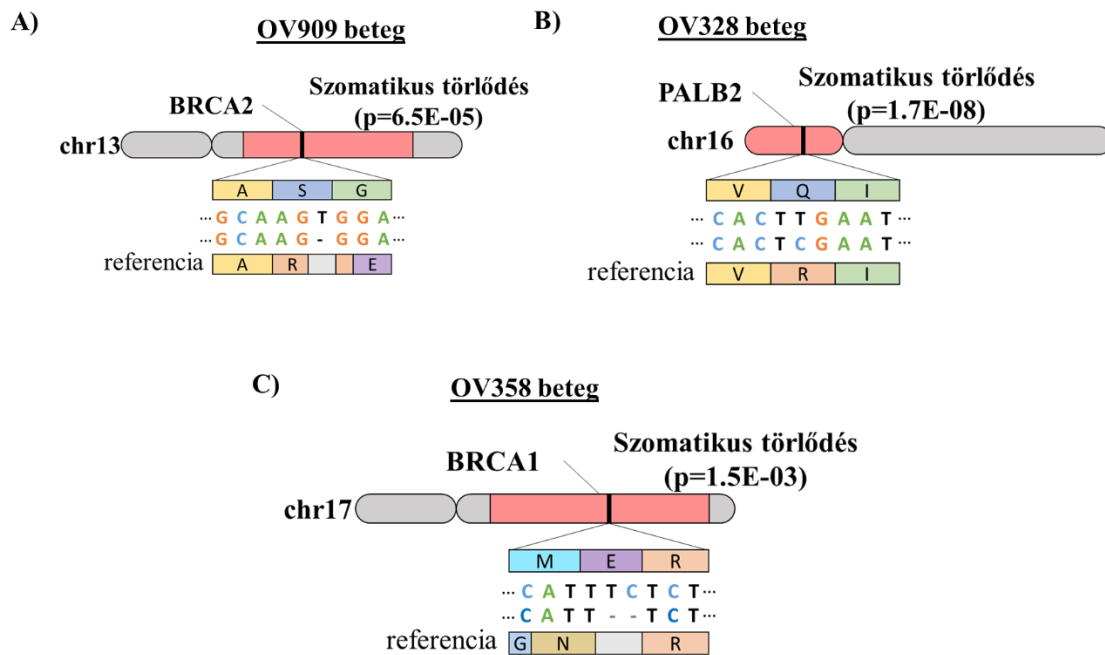
5.2. A petefészek tumorok multirégiós szekvenálása

A vizsgálat során öt petefészek tumorból végeztünk újgenerációs teljes exom szekvenálást a szekvenált tumor minta méret és a detektált szomatikus mutációk számának összefüggésének meghatározására.

Mindegyik betegnél meghatároztam az öröklött mutációkat a normál (vér) mintából végzett szekvenálások eredményeiből. Az azonosított mutációkat funkcionálisan annotáltam az *SNPeff* programmal. A populációs SNP-eket a dbSNP adatbázissal határoztam meg, a ClinVar adatbázis segítségével patogenitással összefüggő variánsokat kerestem. Az öröklött mutációk közt 7 variáns rendelkezett bizonyított patogén státusszal, ezek a BRCA1, a BRCA2, a CFH, az ABCC6, az MPO, az NTRK1 és a NUP93 géneket érintették.

5.2.1. Homológ rekombinációban fontos szerepű gének mutációi

Három mintában azonosítottam a DNS kettős-szál törésének javításában kulcsszerepű génekben öröklött mutációkat. A 909-es betegben a BRCA2 génben egy olvasási keretet eltoló egybázisos törlődést találtam (21/A ábra), ami a fehérje aminosav szekvenciáját teljesen megváltoztatta. A 328-as mintában egy glutaminból arginin aminosav cserét okoz mutációt azonosítottam a PALB2 génben (21/B ábra). A PALB2 génben több patogén variáns található a ClinVar adatbázisban, így feltételezhetően az azonosított mutáció hozzájárulhat a tumor genetikai instabilitásához. A 358-as mintában a BRCA1 génben találtam egy kétbázisos olvasási keret eltolódást okozó törlődést (21/C ábra). A három öröklött mutációt vizuálisan igazoltam az IGV (Integrated Genomic Viewer [148]) program segítségével. Mindhárom mutációnál törlődött a vad típusú allél a tumorban, esetükben szignifikánsan növekedett a tumor mintákban mért mutációk frekvenciája a normál mintákhoz képest (BRCA1: $p=1.5E-03$; BRCA2: $p=6.5E-05$; PALB2: $p=1.7E-08$) Fisher-teszt alapján.

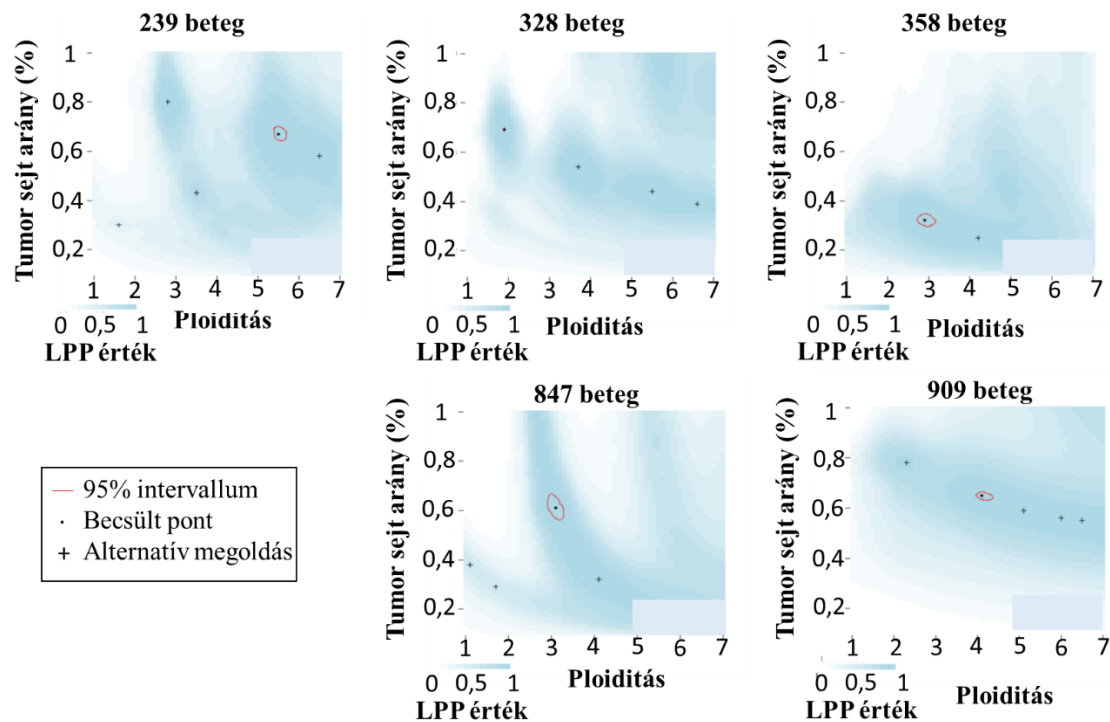


21. ábra. Azonosított öröklött mutációk homológ rekombinációban (HR) résztvevő génekben. Minden esetben szomatikusan törlődött a vad allél a tumorban. A) A 909-es betegben egy BRCA2 kereteltolódásos deléció volt található. B) A 328-as betegben egy öröklött PALB2 mutáció volt található. C) A 358-as betegben a BRCA1 génben volt található egy kereteltolódást okozó mutáció.

5.2.2. Kópiaszám-változás elemzése és a tumor-arány számítása

Tumorokban gyakoriak a genom instabilitásából származó gén, illetve kromoszóma kópiaszám változások. A megfelelő becslésükhöz fontos paraméter a tumorsejtarányok pontos számítása.

A kópiaszám-változások számításának első lépése a tumorsejtarányok és a tumor-ploiditás értékek meghatározása volt a *sequenza* programmal. Az illesztett adatok feldolgozása során a program optimális értékeket becsült mind a két paraméterre, melyekkel elvégezte az automatizált teljes feldolgozásokat. A tumorsejtarányok, a ploiditás értékek és az utólagos valószínűségük (LPP érték) függvényében a program optimális mellett szub-optimális értékeket is megjelölt az LPP értékek lokális maximumai alapján. (22. ábra). A tumorsejtarányok a tumorokban lévő normál szövet kontaminációk arányáról adott információt. A ploiditás értékek viszont a tumorok genom másolatainak számát adta meg (diploid genom esetén kettő), ami hibás osztódás során változhat.



22. ábra. A tumorsejtarány (cellularitás) és a tumor ploiditás utólagos valószínűségi diagramok a betegek biopszia régiójáról. Minden diagramban megtalálható az optimális (fekete pont) és szub-optimális érték (kereszt).

A megfelelő tumorsejtarány és ploiditás értékelése során azon értékeket választottam, ahol a ploiditás a leghasonlóbb volt egy beteg három régiója között. Azon esetekben, ahol egy ploiditás értékhez több tumorsejtarány érték társult, ott a többi régióhoz leghasonlóbb tumor arány értéket választottam. A választott értékeket az 8. táblázatban tüntetem fel.

8. táblázat. Az egyes beteg mintáinak választott tumor arány és ploidiás értékei.

	Biopszia		Lokális		Globális	
Minta	Tum. arány	Ploiditás	Tum. arány	Ploiditás	Tum. arány	Ploiditás
239	80,00	2,80	0,76	2,80	86,00	2,70
328	69,00	1,90	76,00	1,90	82,00	1,90
358	32,00	2,90	34,00	3,10	43,00	3,30
847	61,00	3,10	54,00	2,20	18,00	2,20
909	78,00	2,30	80,00	2,30	80,00	2,50

A tumorsejtarány és ploidiás értékek meghatározás után elvégeztem a kópiaszám változás elemzést. A legtöbb kópiaszám változással rendelkező betegben 87 szegmens volt érintett. A legkevesebb eseménnyel rendelkező betegben összesen 47 változást találtam. Az egyes régiók közt az események nagyfokú homogenitást mutattak, legalább 95,7% fedték le egymást a hosszuk több mint 95%-ában. Az egyes tumorokban azonosított kópiaszám változások számát a 9. táblázatban foglaltam össze.

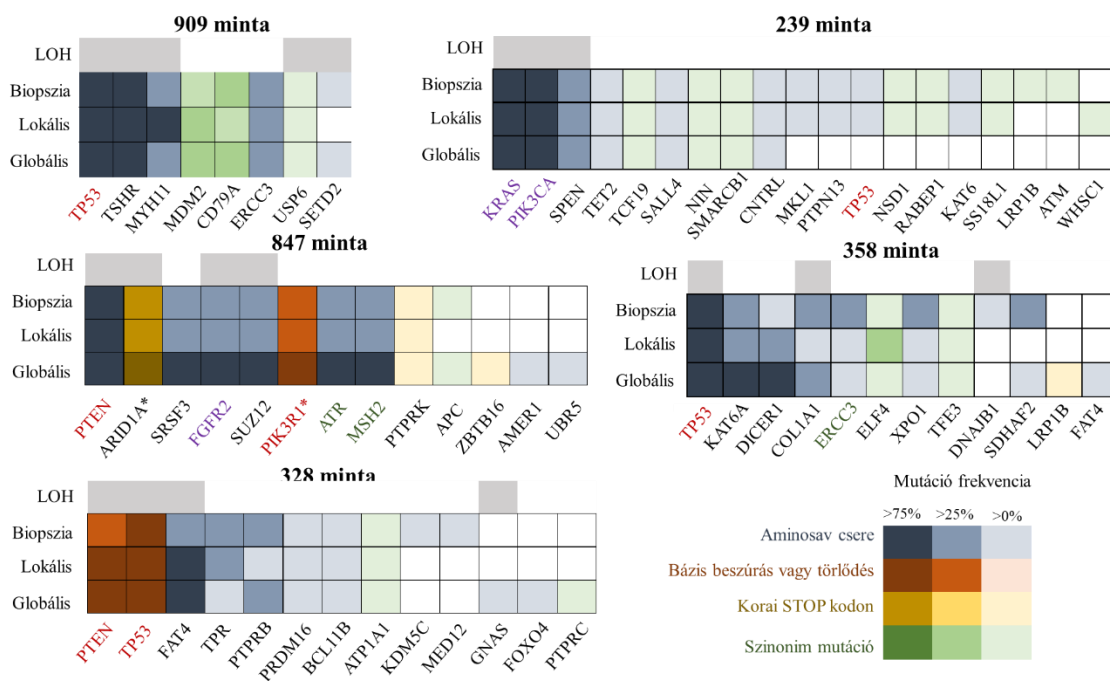
9. táblázat. A kópiaszám változás események száma, valamint a százalékos átfedésük. Látható, hogy a kópiaszám változások legalább 93%-ban egyeztek meg az egyes minták között.

	Kópia szám változás (db)	Közös (%)		
Minta		Biopszia	Lokális	Globális
239	55	100	98,2	98,2
328	65	98,5	98,5	98,5
358	47	95,7	97,9	100
847	47	100	95,7	93,6
909	87	96,6	96,6	98,9

5.2.3. Szomatikus mutációk azonosítása

Szomatikus mutációk keresése során olyan bázissorrend változásokat kerestem, melyek jelen voltak a tumorban, viszont a normál szövetből hiányoztak. A szomatikus mutációk azonosítását a *mutect2* illetve az *mcaller* programokkal végeztem a 14. ábra alapján.

Az öt betegben azonosított szomatikus mutációk száma lényegesen eltért egymástól. A 23. ábrán láthatók a COSMIC tumor konszenzus géneket érintő mutációk. A legtöbb mutáció aminosav cserével járt, a többi mutáció vagy nem okozott aminosav csere, vagy a mutáció romboló jellegű volt.



23. ábra. A tumor mintákban azonosított COSMIC tumor gén konszenzus mutációk. A cellák színe a mutáció típusát jelölik (kék: aminosav csere; narancs: inszerció/delécio; sárga: korai stop kodon beépülése; zöld: csendes mutáció), árnyalatuk a mutáció frekvenciával arányos. A mátrix tetején vannak jelölve a heterozigotáság veszteséssel (LOH) érintett mutációk.

A legtöbb mutációt a 239-es betegben azonosítottam. A mintában kimutattam a leggyakoribb KRAS gén p.G12D aminosav cserével járó aktiváló mutációt, valamint a leggyakoribb PIK3CA gén p.E454K aminosav cserével járó mutációt. Mindkét gént lefedte a tumorban egy gén törlődés.

A 847-es mintában azonosított mutációk közül négy a PI3K növekedési jelpályán helyezkedett el. A legmagasabb allélfrekvenciával rendelkező mutáció a PTEN tumorszupresszor génben volt. Emellett a PI3K útvonalat serkentő FGFR2 receptor génben leggyakrabban előforduló mutációt azonosítottam. Az utolsó két mutáció PI3K útvonal aktivitását mérséklő PIK3R1 génben lettek azonosítva.

Az utolsó három betegben a homológ rekombináció hibajavító útvonalban azonosítottam öröklött heterozigóta mutációkat, esetükben szomatikus heterozigótaság vesztés által törölték a vad allélok. A 358-as betegben a legmagasabb allél frekvenciával rendelkező szomatikus mutáció aminosav cserét okozott a TP53 génben. A TP53 génnél egy törlődést is azonosítottam. A mintában azonosítható volt egy alacsony frekvenciájú mutáció az egyszálú hibajavításban fontos szerepű ERCC3 génben. A 328-as betegben a TP53 és a PTEN tumorszuppresszor génekben találtam szomatikus mutációkat. Mindkét génnél törlődött az a vad allél. A 909-es mintában a TP53 illetve a TSHR génekben találtam szomatikus mutációkat.

5.2.4. Egyes tumor régiókban azonosított mutációk számának összehasonlítása

A mutációk az egyes régiók között 15,5 – 93,8% intervallumban egyeztek meg. Az egyes mintákban azonosított mutációk számát és közös mutációk arányát a 10. táblázat tartalmazza.

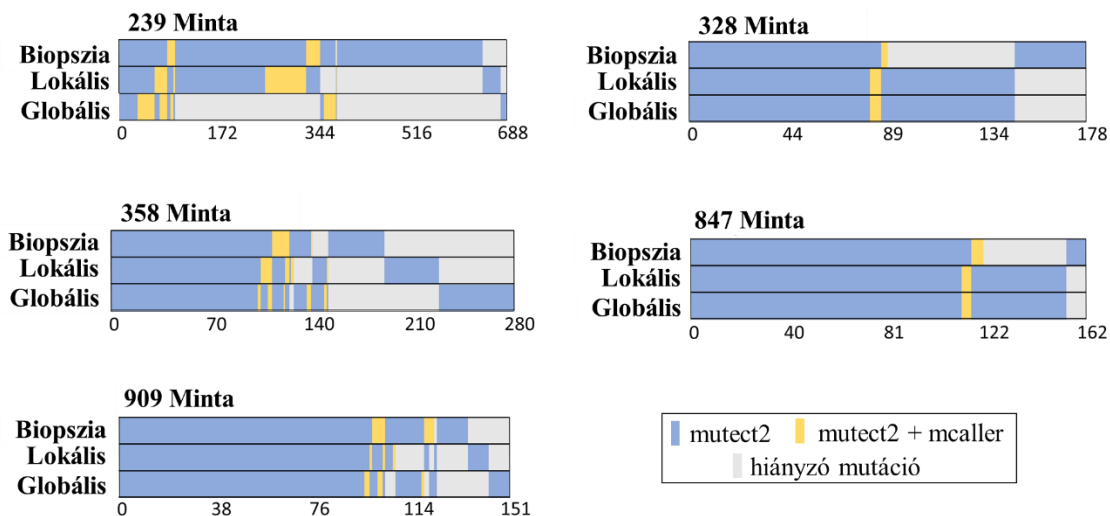
10. táblázat. Az egyes tumor régiókban azonosított mutációk száma. Egy betegben a mutációk legalább 15,5%-ban egyeztek meg az egyes régiókban.

	Biopszia		Lokális		Globális	
Minta	Össz. mutáció	Közös (%)	Össz. mutáció	Közös (%)	Össz. mutáció	Közös (%)
239	644	15,5	392	25,5	140	71,4
328	123	72,4	146	61,0	146	61,0
358	179	69,3	176	70,5	200	62,0
847	128	93,8	154	77,9	154	77,9
909	134	76,9	118	87,3	125	82,4

A 239-es beteg biopszia mintájában volt a legtöbb mutáció azonosítható (688 db). Esetében a minta méret növekedésével csökkent a detektálható mutációk száma. A globális mintában már csak 140 mutációt azonosítottam, ebből 71,4% volt megtalálható a biopszia és lokális mintákban is.

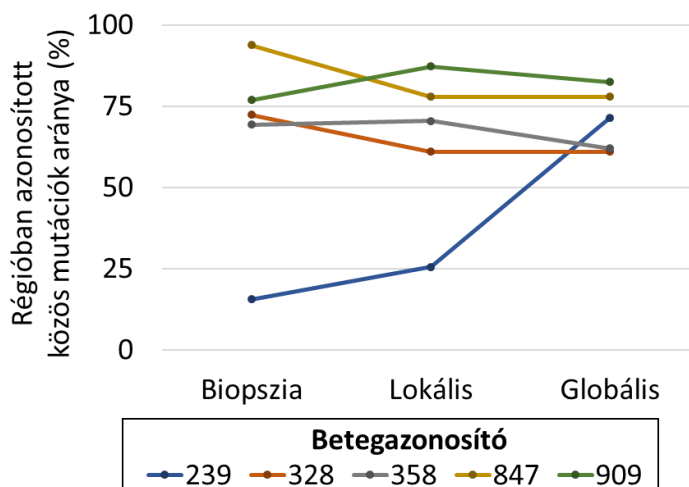
A 328-as és 847-es betegek mutatták a legnagyobb hasonlóságot a lokális és a globális minták között. Ezen betegeknél a lokális és a globális mintákban ugyanazokat a mutációkat detektáltam, csak a frekvenciájuk különbözött. Mindkét betegnél a biopszia mintában volt a legkevesebb mutáció azonosítható.

A 358-as és a 909-es betegeknel a biopszia és a globális minták mutatták a legnagyobb hasonlóságot.



24. ábra. A szomatikus mutációk mutációs mátrixa. Az ábrában a sorok jelölik az egyes régiókat, az oszlopok a mutációkat. A mutációk három besorolást kaptak: *mutect2*-vel azonosított mutációk (kék), *mutect2*-vel nem, de *mcaller*-el igazolt mutációk (sárga) és hiányzó mutáció (szürke).

A mutációs mátrixokban látható (24. ábra), hogy az egyes minták közt lényegesen különböztek a régiókban talált közös mutációk aránya. A 239-es beteg biopszia mintájában a mutációk csak 15,5%-a volt megtalálható a lokális és globális mintákban. A többi betegben ez az arány 69,3%-93,8% között volt. A lokális minták esetén is kiugrott a 239-es minta, viszont a globális mintáknál már nem (25. ábra).



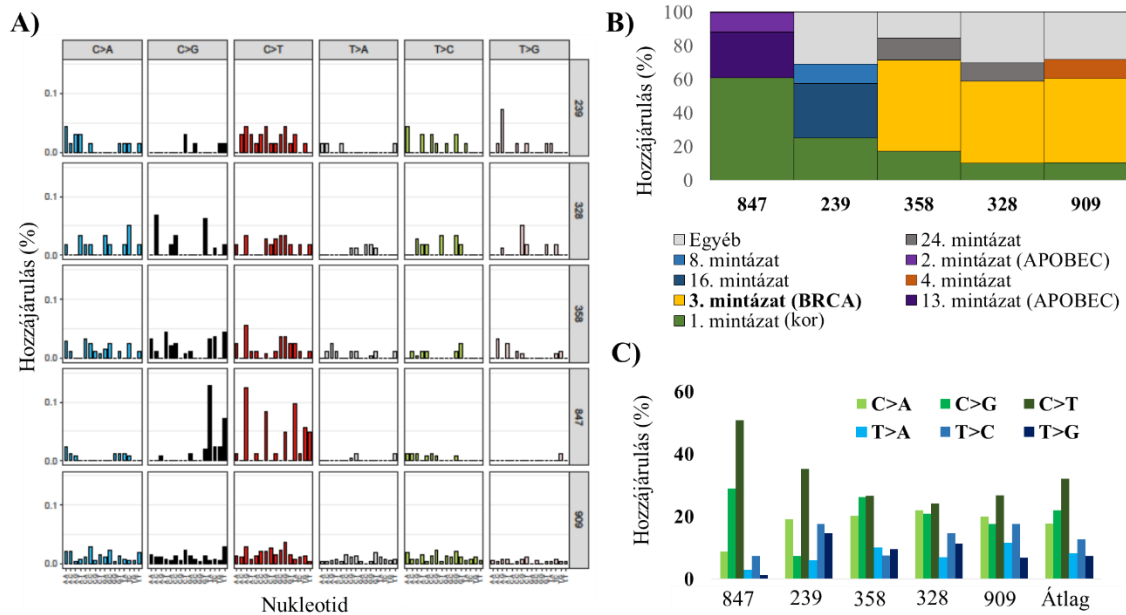
25. ábra. Egyes mintákban azonosított közös mutációk aránya.

5.2.5. Szomatikus mutáció mintázatok

Minden betegnél kiszámoltam a mutációs mintázatokat a *MutationSignatures* program segítségével. Ennek során a mutációkat egy olyan hisztogramban ábrázoltam, mely tri-nukleotid kontextusban jeleníti meg az egy bázis cserével járó mutációk arányát (26/A ábra). Az ábrán látható, hogy a mutációs mintázatok különböztek az egyes minták között. Mivel az újgenerációs szekvenálás során nem lehetett eldönteni a leolvasások irányát, emiatt a szubsztitúciók 12 kombinációját tükrözéssel hat lehetőségre szűkítettem.

Az egyes mutációs mintázatok összetételét a 26/B ábrán lettek feltüntetve. Minden mintában azonosítható volt a korral járó mintázat (zöld), ami az 5-metilcitozin spontán dezaminációjából származik. A 847-as betegnél az APOBEC citidin-dezamináz hibával járó mutációs mintázatokat határoztam meg (sötét és világos lila). Az APOBEC mintázatoknál magasabb a C>T és a C>G bázis változások aránya (26/C ábra). A 239-es betegnél két ismeretlen eredetű mintázatot találtam.

A 328-as, 358-as és 909-es betegeknél a BRCA hibás működésével asszociált mintázatot határoztam meg (sárga). A mintázat a DNS kettős szál törések javításának hiányával van összefüggésben, ami kis és nagy DNS törlődésekkel, valamint genom átrendeződésekkel jár. A három betegnél a bázisváltozás típusai sokkal egyenletesebben oszlottak meg a 847-es és 239-es mintákhoz képest (26/C ábra).



26. ábra. Szomatikus mutációk jellemzése. A) Az egyes betegek mutációs mintázata. B) A mutációs mintázatok összetétele irodalmi adatok alapján [26]. Minden betegben azonosítható volt a korrall összefüggő 1-es mintázat (zöld). A 847-es betegben látható volt az APOBEC citidin-dezamináz hibával társuló 2-es, illetve 13-as mintázatok (lila). A 239-es betegben két ismeretlen eredetű mintázat volt azonosítható (8-as és 16-os mintázatok). Az utolsó három betegben azonosítható volt a BRCA hibával összefüggő 3-as mintázat (sárga). A 358 és 328 betegekben azonosítható volt az aflatoxin kitétséghez társított 24-es mintázat (sötétszürke). A 909-es betegben azonosítható volt a dohányzással összefüggésbe hozott 4-es mintázat (narancs). C) Szubsztitúciók százalékos ábrázolása.

5.3. Túlélés elemzés mutációhoz kapcsolt génkifejeződés változással emlőtumoros betegekben

Túlélés elemzés során a betegek csoportra bontása gyakran egy génmutáció státusza, vagy kifejeződési szintje alapján történik. A TCGA adatbázist felhasználva sikerült bemutatnom, hogy az AR/FOXA1 útvonalat érintő génmutációk által létrejött BRCA hibajavítás deficiens tripla negatív emlő tumorok érzékenyebbek lehetnek bizonyos kemoterápiás kezelésekre. Ezek független adatbázisok, valamint minták felhasználásával is validálva lettek [149]. A génmutációk együttes vizsgálata útvonal szinten adhat információt a változások hatásáról, a megközelítés hátránya, hogy más változás típusokat nem vesz figyelembe. Emiatt munkám során kifejlesztettem a G-2-O

(Genotype2Outcome) rendszert, amely összekapcsolja a genetikai mutációkat génkifejeződési változásokkal (mint egy kifejeződési ujjlenyomat), és túlélés elemzést végez egy független génkifejeződési adatsoron [147]. Az elemzőrendszer szabadon felhasználható, online elérhető a www.g-2-o.com címen.

Munkám során az emlőrákban öt leggyakrabban mutálódó gén alapján végeztem túlélés elemzést [58]. Az egyes génekben azonosított mutációk összegét a 11. táblázatba foglaltam össze. A táblázatban feltüntettem a *cbioportal.org* honlapról letöltött METABRIC [150] független adatbázisban lévő adatokat is. A legtöbb mutáció a PIK3CA génben volt található, a betegek közel 50%-ában volt mutáció azonosítható mind a két adatbázisban.

11. táblázat. Az emlődaganatokban leggyakrabban mutálódó génekben azonosított mutációk száma a G2O illetve METABRIC adatokban.

	Genotype2Outcome		MetabRIC (cbioportal.org)	
	Mutációval	Vad típusú	Mutációval	Vad típusú
PIK3CA	306	672	840	1026
TP53	258	720	664	1202
GATA3	19	959	316	1550
MAP3K1	51	927	206	1660
MAP2K4	35	943	107	1759
Össz. beteg		978		1866

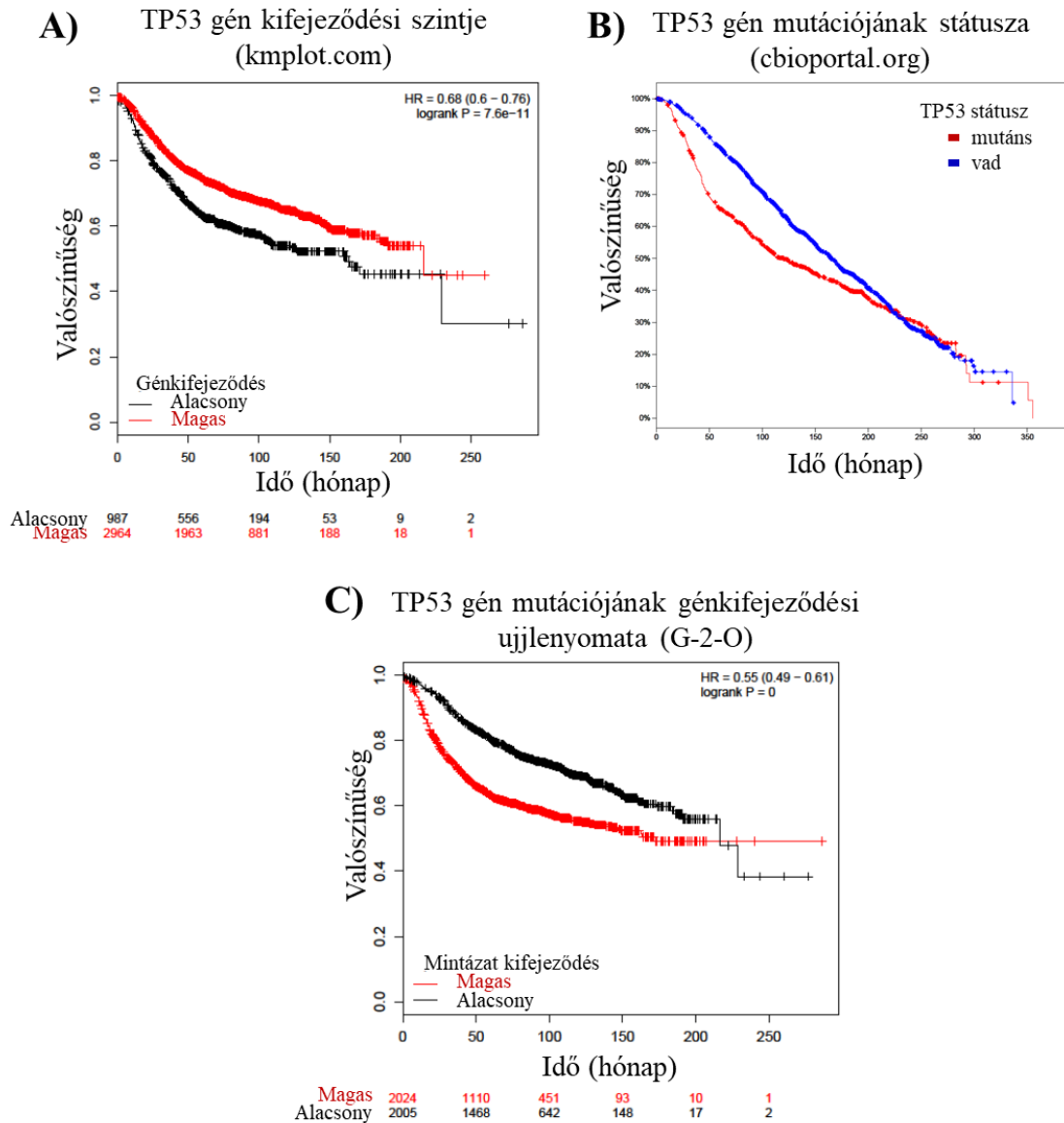
Összehasonlításként a túlélés elemzéseket több rendszerrel is elvégeztem. A *kmplot.com* honlappal a kiválasztott gén kifejeződési szintje alapján végeztem a túlélés elemzést. Minden esetben az optimális vágópont („best cutoff”) automatikus kiválasztásával történt az elemzés, a többi paraméter alapbeállításra volt állítva. A *cbioportal.org* honlapon lévő METABRIC adatok felhasználásával a mutációk státusza (jelenléte) alapján végeztem a túlélés elemzés. A G-2-O elemzőrendszer segítségével a kiválasztott gén mutációjához köthető génkifejeződés változás (mint egy ujjlenyomat) felhasználásával végeztem a túlélés elemzés. Az elemzés folyamata a 15. ábrán látható. Az egyes elemzőrendszerekkel kapott eredményeket a 12. táblázatban foglaltam össze.

12. táblázat. Az emlődaganatokban gyakran mutálódó génekkel végzett túlélés elemzés eredményei.

	Metabric (Mutáció státusz, cbioportal.org)		kmplot.com (kifejeződés)		Genotype2Outcome (kombinált mintázat)	
	p-érték	Rossz prognózis	p-érték	Hazard rate (kockázati ráta)	p-érték	Hazard rate (kockázati ráta)
PIK3CA	NS		2,1e-09	1,39(1,25–1,55)	1,1E-06	1,6(1,4-1,8)
TP53	6,5E-05	mutáció	7,6e-11	0,68(0,6–0,76)	<E-16	0,55(0,59-0,62)
GATA3	1,2E-03	vad	1,3e-08	0,71(0,63–0,8)	<E-16	1,6(1,4-1,8)
MAP3K1	NS		2,9e-14	0,55(0,47–0,64)	<E-16	1,6(1,5-1,8)
CDH1	NS		1,5e-03	1,2(1,1–1,3)	3,9E-09	1,4(1,2-1,5)

A PIK3CA onkogén mutációs állapota nem befolyásolta szignifikánsan a betegek túlélését, viszont magas kifejeződési szintje lényegesen rosszabb prognózissal járt. A PIK3CA génmutációhoz kötött génkifejeződési ujjlenyomat magas kifejeződése jobb túléléssel társult.

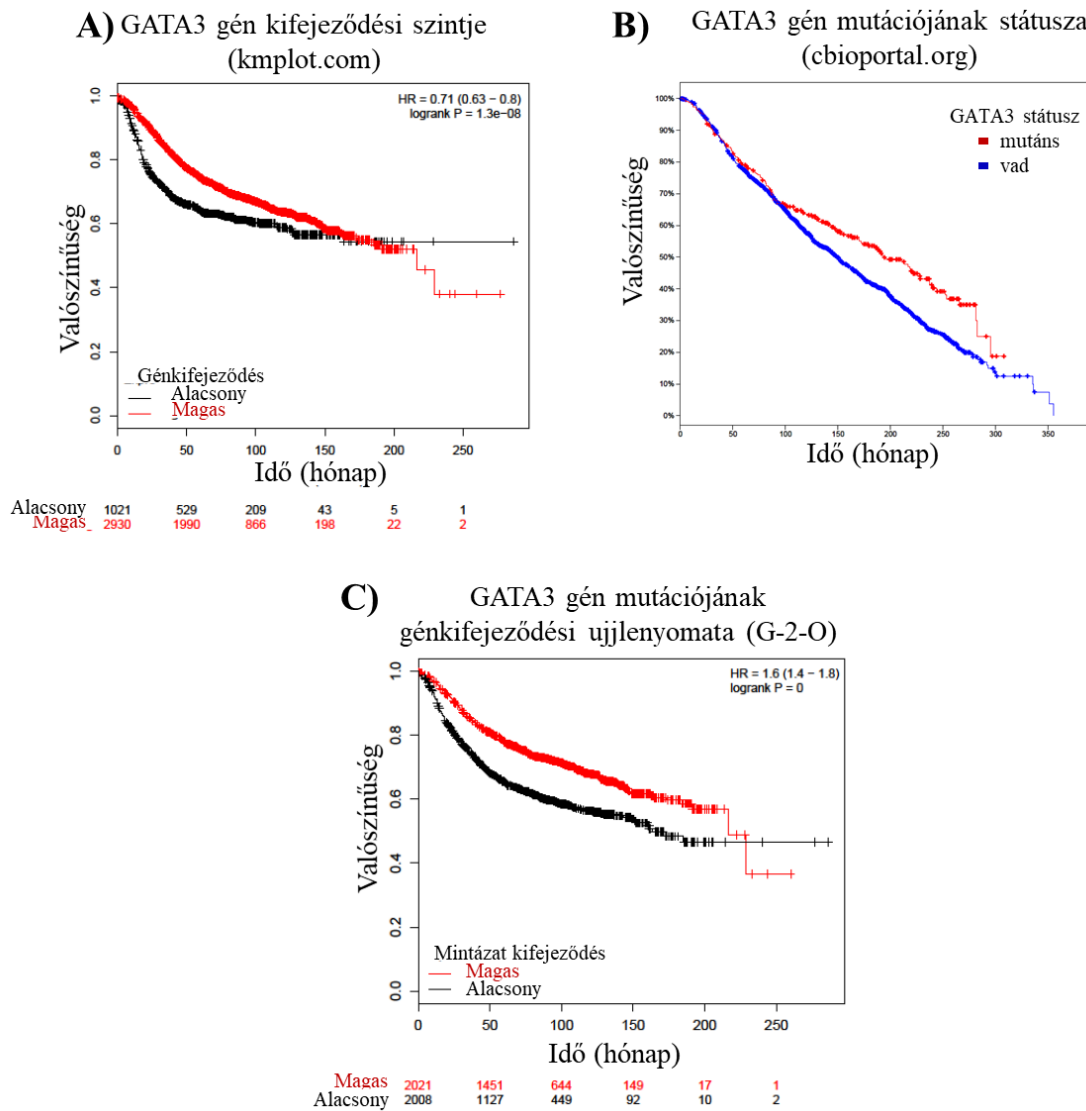
A TP53 gén esetén minden elemzőrendszerrel szignifikáns eredményt kaptam. Az eredményekből látható, hogy a TP53 gén magas kifejeződése jobb túlélési idővel társult (27/A ábra), mutációjakor viszont rosszabb volt a betegek túlélési ideje (27/B ábra). A megfigyelés önmagában nem meglepő, mivel a TP53 egy tumorszupresszor gén. A TP53 gén mutációjához kötött génkifejeződési ujjlenyomat magas kifejeződésekor lényegesen rosszabb volt a betegek túlélési ideje (27/C ábra).



27. ábra. A TP53 gén A) kifejeződése, B) mutációja, és C) mutációhoz kötött génkifejeződés mintázata alapján végzett túlélés elemzés eredményei.

A GATA3 gén esetében az eredmények valamelyest ellentétesnek tűnhetnek. Magas kifejeződése jobb prognózissal járt (28/A ábra), viszont mutációja esetén is jobb volt a prognózis (28/B ábra). Ezzel szemben a génmutációk génkifejeződés ujjlenyomatának alacsony kifejezése járt rosszabb prognózissal (28/C ábra). A METABRIC adatbázisban 52 szubsztitúció, 5 kereteltolódással nem járó bázis törlődés és beszúrás, valamint 193 törlő mutáció volt azonosítható. Ezek mellett 98 betegnél génamplifikációt is találtak. Az eredmények dualítására egy magyarázat, hogy a GATA3 transzkripciós faktor működése fontos a tumor fejlődésében. Funkciójának fokozását a

gén amplifikációjával érheti el, valamint törlő mutációk hatására növelheti a fehérje fél-élet idejét is [151].



28. ábra. A GATA3 gén A) kifejeződése, B) mutációja, és C) mutációhoz kötött génkifejeződés mintázata alapján végzett túlélés elemzés eredményei.

A MAP3K1 gén magas kifejeződése jobb túlélési idővel társult, viszont mutáció státusza nem volt összefüggésbe a betegek túlélésével. A mutációs ujjlenyomat magas kifejeződése jobb túléléssel társult. A CDH1 gén magas kifejeződése rosszabb túlélési idővel társult, a génmutációk kifejeződési ujjlenyomatának magas kifejeződése esetén jobb volt a prognózis.

5.3.1. Túlélés elemzés az emlőtumor szubtypusokban

Az emlőtumorsejtet három fő altípusba lehet sorolni a Szt. Gallen-i osztályozás szerint. Ezek az 1) Luminal A és B altípusok: ösztrogén receptor pozitív, HER2 negatív és alacsony vagy magas KI67 indexű tumorok; a 2) HER2 pozitív altípus: HER2 receptor pozitív tumorok; és a 3) tripla negatív altípus: az ösztrogén, HER2 és progeszteron receptorok státusza negatív. Az altípusok nem csak a receptorok státuszában, de a mutációk szintjén is különböznek. Az egyes altípusokban legtöbb mutációval rendelkező géneket a 13. táblázatban foglaltam össze. A Luminal A és B altípusoknál a gének megegyeztek, így összevontam. A táblázatban „*” jellel jelöltem a G-2-O adatbázisából hiányzó géneket.

13. táblázat. Az emlőtumorsejt egyes altípusaiban az öt leggyakrabban mutálódó gén mutációinak száma a G-2-O adatbázisában.

Altípus	Gén	Mutációval	Vad
Luminal A/B	PIK3CA	267	442
	TP53	108	601
	CDH1	61	648
	MAP3K1	44	665
	GATA3	17	692
HER2 pozitív	PIK3CA	99	191
	TP53	62	228
	PTPN22	22	268
	CDH1	9	281
	AFF2*	NA	NA
Tripla negatív	TP53	119	90
	PIK3CA	25	184
	RB1	15	194
	MLL3*	NA	NA
	AFF2*	NA	NA

Az összehasonlító elemzéseket csak a G-2-O valamint a *kmplot.com* rendszerekkel végeztem el, mivel esetükben altípusokra lehetett bontani a betegeket. A *cbioportal.org* honlappal csak teljes tumor adatsorok elemzésére van lehetőség. A kapott eredményeket a 14. táblázatban foglaltam össze.

14. táblázat. Az 5 leggyakrabban mutálódó génnel végzett túlélés elemzés eredményei emlőtumorok egyes altípusaiban

Altípus	Gén	Genotype2Outcome (kombinált mintázat)		kmplot.com (kifejeződés)	
		p-érték	Hazard rate (kockázati ráta)	p-érték	Hazard rate (kockázati ráta)
Luminal A	PIK3CA	9E-03	1,3(1,1-1,5)	1,20E-05	1,48(1,24-1,76)
	TP53	<E-16	0,47(0,39-0,56)	3,40E-07	0,64(0,53-0,76)
	MAP3K1	2,9E-03	1,3(1,1-1,5)	4,70E-09	0,48(0,38-0,62)
	GATA3	NS		NS	
	CDH1	NS		1,40E-04	1,41(1,18-1,68)
Luminal B	PIK3CA	NS		4,50E-04	1,42(1,17-1,73)
	TP53	4,5E-02	0,82(0,68-1,0)	2,50E-02	0,8(0,6-0,97)
	GATA3	1,5E-02	1,3(1,0-1,5)	2,00E-04	1,6(1,25-2,05)
	CDH1	2,6E-03	1,3(1,1-1,6)	2,60E-03	1,34(1,11-1,62)
	MAP3K1	4E-03	1,3(1,1-1,6)	NS	
HER2 pozitív	PIK3CA	1,4E-02	0,63(0,43-0,91)	NS	
	TP53	1,1E-02	0,62(0,42-0,9)	1,30E-04	0,48(0,33-0,7)
	PTPN22	NS		1,80E-04	0,43(0,27-0,68)
	CDH1	NS		1,80E-04	2,13(1,42-3,2)
	AFF2*	NS		NS	
Tripla negatív	PIK3CA	NS		4,10E-03	1,46(1,13-1,9)
	TP53	1,2E-02	1,4(1,1-1,8)	1,40E-02	0,71(0,54-0,93)
	MLL3*	NS		NS	
	RB1	NS		NS	
	AFF2	NS		6,00E-04	0,63(0,48-0,82)

Az elemzések során csak a TP53 gén esetén kaptam szignifikáns eredményt mindegyik altípusban a két elemzőrendszerrel. A TP53 génmutációk génkifejeződési ujjlenyomatának magas kifejeződése rosszabb prognózissal járt a tripla negatív altípus kivételével, ahol az ujjlenyomat alacsony kifejeződése társult rosszabb prognózishoz.

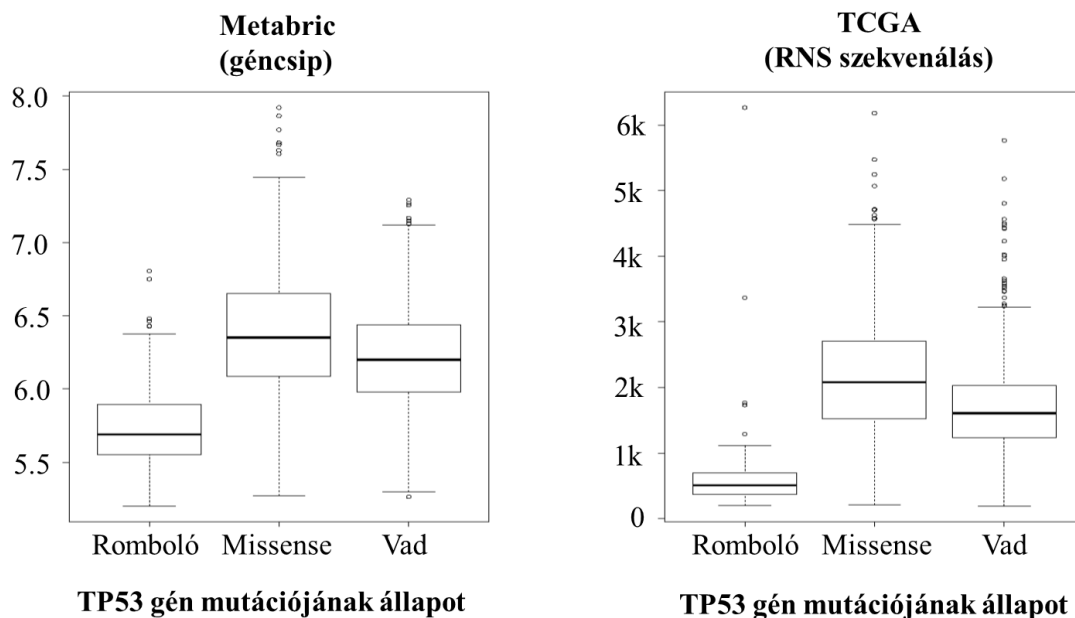
A gének egy részében a mutáció ujjlenyomatának kifejeződése nem társult szignifikáns túlélés különbséghez, viszont génkifejeződési szintje igen. A Luminal A altípusban a CDH1 génnél, Luminal B altípusban a PIK3CA génnél, HER2 pozitív altípusban a PTPN22, a CDH1 és az AFF2 géneknél, valamint tripla negatív altípusban a

PIK3CA és az AFF2 géneknél volt megfigyelhető. Ezzel szemben a Luminal B altípusban a MAP3K1 génmutációk, valamint HER2 pozitív altípusban a PIK3CA génmutációk génkifejeződési ujjlenyomata szignifikáns túlélés különbséghez társult, viszont a gének kifejeződési szintje önmagában nem.

5.3.2. A TP53 génmutációhoz kötött génkifejeződés változás

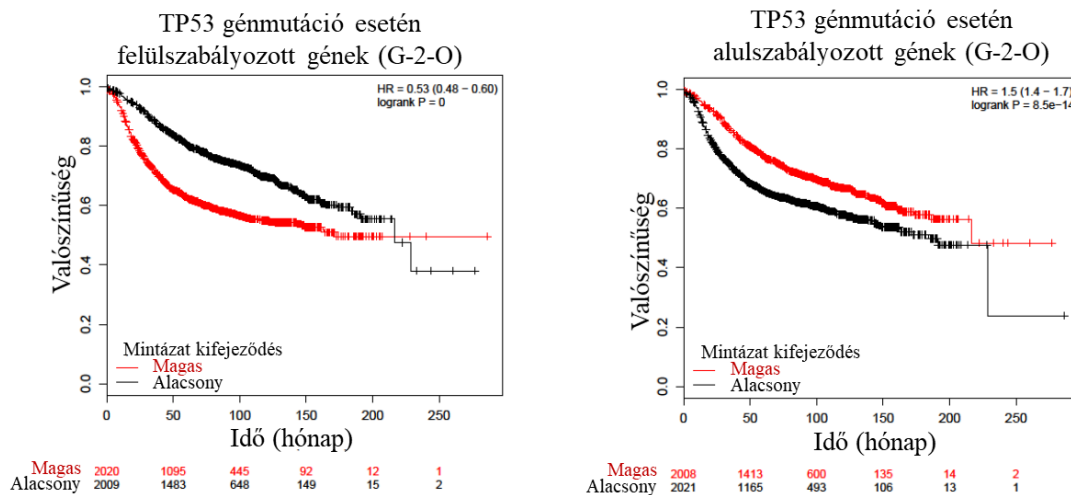
A TP53 génmutációk szinte minden tumor típusban gyakoriak [7]. A mutációs állapotát több tumor típusnál összefüggésbe hozták a betegek prognózisával és onkológiai kezelések hatékonyságával. Munkám célja volt megvizsgálni a TP53 génmutációkhoz asszociált génkifejeződés változást, mely segíthet megérteni jelentőségét a tumor fejlődésében.

A TP53 gént több féle mutáció érintette. Tumorszupresszorként sok olyan mutációt tartalmazott, ami korai STOP kodonokat, valamint bázis beszúrásokkal és törlődésekkel olvasási keret eltolódást okoztak (2. ábra). A G-2-O illetve a METABRIC adatbázisokban látszódott, hogy a mutáció jelenlététől és jellegétől (romboló, szubsztitúció, vad) függően különbözött a gén kifejeződési szintje (29. ábra). Azt tapasztaltam, hogy a romboló mutációknál lényegesen alacsonyabb volt a TP53 gén kifejeződési szintje.



29. ábra. A TP53 gén kifejeződési szintje a mutáció típusának függvényében TCGA és METABRIC adatsorokban.

A G-2-O elemzés alapján látható volt, hogy a mutációjával összefüggő génkifejeződés változás összefüggésbe hozható a túlélési idővel. A felülszabályozott gének magas kifejeződése rossz prognózissal járt. Az alulszabályozott gének alacsony kifejeződési szintje szintén rossz prognózissal járt (30. ábra).



30. ábra. A TP53 génmutáció esetén felül-, illetve alulregulált génekkel végzett túlélés elemzés.

5.3.3. Gén ontológia elemzés a TP53 gén mutációjával asszociált génekkel emlő és NSCLC betegeknél

A DAVID honlap (david.ncifcrf.gov [152]) segítségével gén ontológia elemzést végeztem a 100 legjobban felülszabályozott gén felhasználásával. Az elemzés segített egy várható értéknél gyakrabban érintett útvonalaknak az azonosításában. Az alulszabályozott génekkel nem kaptam szignifikáns eredményt, ahol a korrigált p-érték alacsonyabb lett volna $p < 0,05$ értéknél (csak ha > 500 növeltem a génlista hosszát). Az elemzést elvégeztem emlőtumoros betegek adatainak, valamint összehasonlításként nem-kissejtes tüdőrákos (NSCLC) betegek adatait felhasználásával is [153]. A top 100 felül- illetve alulszabályozott génnel együttesen végzett génontológia elemzés során a glikolízissel összefüggő útvonalakat is azonosítottam. Ezen változásokat sejtvonalas kísérletekkel is validáltuk [154].

A felülszabályozott génekkel emlőtumorsebetegek esetén az elemzés 5 útvonalat azonosított $p < 0,05$ korrigált p-értékkel. Ezeket a 15. táblázatban foglaltam össze.

15. táblázat. A TP53 génmutációval rendelkező emlőtumoros betegekben felülszabályozott génekkel végzett gén ontológiai elemzés eredménye.

Útvonal	Útvonal azonosító (KEGG)	Gének (db)	p-érték	korrigált p-érték
Sejt ciklus	hsa04110	23	1,3E-18	1,3E-16
Oocita meiózis	hsa04114	9	1,7E-04	8,4E-03
p53 jelátviteli útvonal	hsa04115	7	3,9E-04	1,3E-02
Progeszteron mediált oocita érés	hsa04914	7	1,5E-03	3,9E-02
DNS replikáció	hsa03030	5	1,7E-03	3,4E-02

A sejtciklus útvonalban 23 gént azonosítottam, melyeknél magasabb volt a kifejeződési szint a TP53 génmutációval rendelkező betegekben.

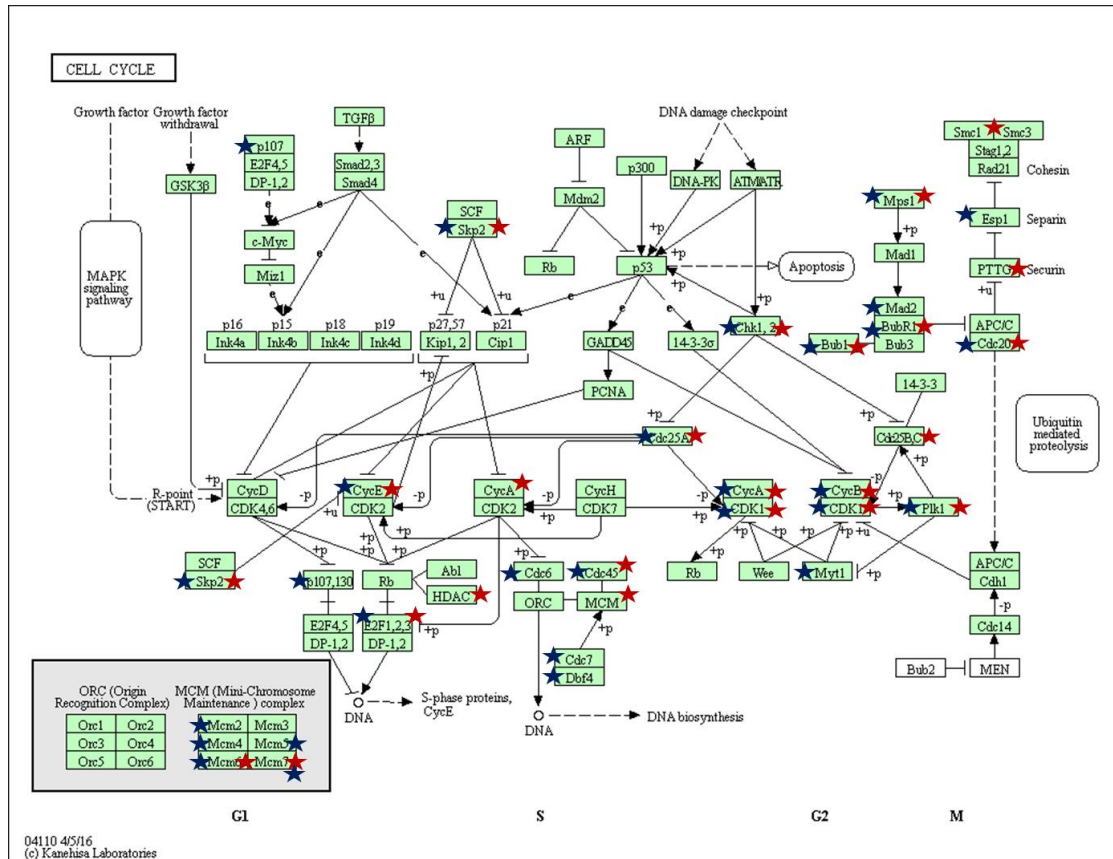
NSCLC-ben a TP53 génmutációjához asszociált ujjlenyomattal végzett gén ontológiai elemzés során 8 útvonalat azonosítottam (16. táblázat). Legkisebb p-értékkel rendelkező útvonal szintén a sejtciklus volt, ez esetben 28 gén kapcsolódott az útvonalhoz. Ezen kívül több hibajavító útvonal volt érintett.

16. táblázat. A TP53 génmutációval rendelkező nem-kissejtes tüdődagamos betegekben felülszabályozott génekkel végzett gén ontológiai elemzés eredménye.

Útvonal	Útvonal azonosító (KEGG)	Gének (db)	p-érték	korrigált p-érték
Sejt ciklus	hsa04110	28	9,7E-26	1,1E-23
DNS replikáció	hsa03030	11	3,5E-11	3,8E-09
Oocita meiózis	hsa04114	11	2,5E-06	2,7E-04
p53 jelátviteli útvonal	hsa04115	9	3,7E-06	4,1E-04
Fanconi anémia útvonal	hsa03460	8	7,9E-06	8,6E-04
Homológ rekombináció útvonal	hsa03440	6	4,5E-05	4,9E-03
Progeszteron mediált oocita érés	hsa04914	8	2,0E-04	2,2E-02
MMR útvonal	hsa03430	5	2,6E-04	2,8E-02

A honlap segítségével lehetőség volt a KEGG útvonalra rávetíteni az azonosított géneket. A sejtciklus útvonalban kapcsolódó géneket a 31. ábrán mutatom be. Az ábrában piros csillagok jelölik az emlőtumorokban, kék csillagok az NSCLC-ben azon géneket, melyek magasabb kifejeződéssel jártak a TP53 génmutációval rendelkező betegekben. Látható, hogy a sejtciklus minden fázisában voltak felül-szabályozott gének, amik

feltételezhetően serkentik a sejtsztódást. Ez összhangban van a KI67 proliferációs marker génkifejeződésének eredményével, ahol a TP53 gén mutáns beteg csoportban körülbelül 2-szeres kifejeződési szintet mutatott az emlőtumorokban és az NSCLC daganatokban is.



31. ábra. A TP53 génmutációval rendelkező betegek esetén a sejtciklus útvonal [155] magas kifejeződést mutató génei emlő (kék) és nem-kissejtes tüdő (piros) daganatokban.

5.3.4. Túlélés-elemzés sejtciklus szabályozással összefüggő génekkel

A génontológia elemzés során 23 olyan gén kifejeződési szintje emelkedett meg az emlőtumorok TP53 génmutációjakor, melyek sejtciklus szabályozásban vesznek részt. Ezen gének kifejeződési szintje alapján végzett túlélés elemzések során szinte minden esetben rosszabb prognózissal járt a magas kifejeződési szint (17. táblázat). Kivételek a CHEK2 és E2F gének, esetükben nem volt szignifikáns különbség a túlélési időben, illetve az RBL1 gén magas kifejeződési szintjével jobb prognózis társult. Az irodalomban

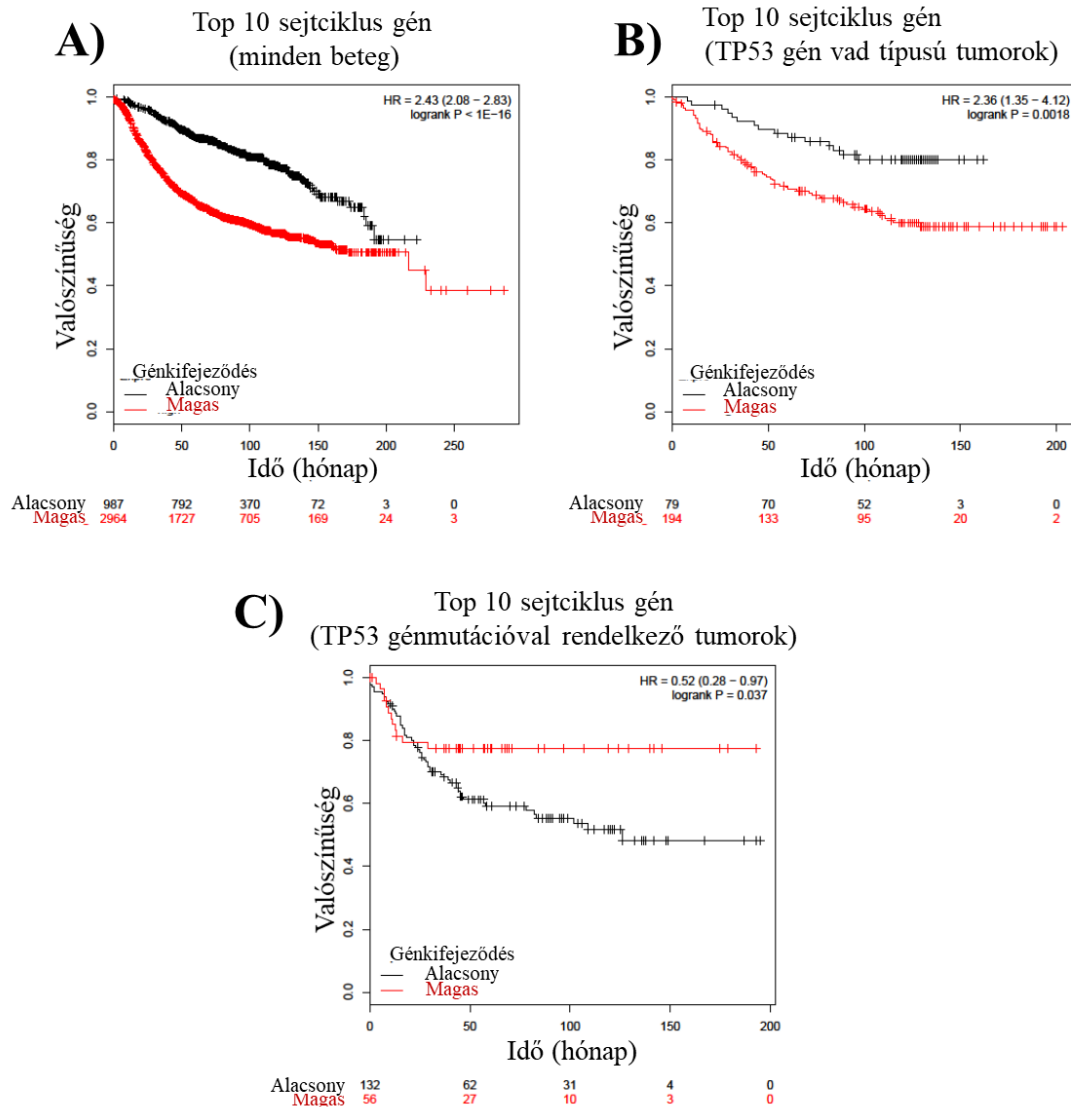
17 génről találtam olyan publikációt, amely fő célja az adott gén kifejeződési szint kapcsolatának vizsgálata emlőtumoros betegek túlélésével.

17. táblázat. A túlélés-elemzések eredménye sejtciklus szabályozó génekkel emlőtumoros betegekben.

Gén	HR	p-érték	Publikáció
CCNB2	2 (1,79-2,23)	1,0E-16	[156]
CDC20	1,91 (1,71-2,13)	1,0E-16	[157]
TTK	1,9 (1,7-2-12)	1,0E-16	[158]
MAD2L1	1,86 (1,67-2,08)	1,0E-16	[159]
CCNA2	1,84 (1,64-2,05)	1,0E-16	[160]
BUB1	1,79 (1,6-2)	1,0E-16	[161]
BUB1B	1,79 (1,61-2)	1,0E-16	[161]
CDC25A	1,76 (1,57-1,96)	1,0E-16	[162]
SKP2	1,75 (1,57-1,96)	1,0E-16	[163]
DBF4	1,63 (1,46-1,83)	1,0E-16	[164]
CDC19	1,53 (1,37-1,7)	2,9E-14	[165]
CDC6	1,5 (1,35-1,68)	2,5E-13	[166]
CCNE1	1,48 (1,33-1,66)	1,1E-12	[167]
PLK1	1,42 (1,27-1,59)	2,5E-10	[168]
MCM7	1,39 (1,25-1,55)	2,9E-09	
CHEK1	1,27 (1,09-1,48)	2,3E-03	[169]
CDC54	1,26 (1,13-1,41)	2,4E-05	
ESPL1	1,25 (1,12-1,4)	4,4E-05	
MCM6	1,21 (1,04-1,41)	1,6E-02	
CDC7L1	1,14 (1,02-1,27)	1,9E-02	[164]
E2F2	1,13 (0,97-1,32)	0,12	[169]
CHEK2	0,96 (0,86-1,07)	0,46	
RBL1	0,78 (0,67-0,91)	1,9E-03	

A *kmplot.com* segítségével lehetőség volt több génnel együttesen túlélés elemzést végezni. Ebben az esetben a bemeneti gének kifejeződésének átlaga alapján végeztem a betegek szétválasztása. A 17. táblázatban lévő első 10 génnel végzett túlélés elemzésből kapott eredményekből látható, hogy a magas kifejeződés rossz prognózissal társult (HR=2,43, $p < E-16$, 32/A ábra). Hasonló volt a helyzet, amikor a beteg csoportot

leszűkítettem a TP53 génmutáció nélküli betegekre ($HR=2,36$, $p=1,8E-3$, 32/B ábra). A TP53 génmutációval rendelkező betegeknél viszont fordított eredményt kaptam (32/C ábra), a génnek magas kifejeződése jobb prognózissal járt ($HR=0,52$, $p=3,7E-2$).



32. ábra. A TP53 génmutációra visszavezethető legjobb 10 megváltozott kifejeződéssel rendelkező génnel végzett túlélés elemzés eredménye. A) Minden beteggel, B) a TP53 gén vad típusú tumorokkal, valamint C) a TP53 gén mutáns tumorokkal rendelkező betegekkal.

6. Megbeszélés

6.1. A sejtmozgás *in vitro* követése újgenerációs szekvenálással

Onkológiai klinikai diagnosztikai szekvenálás során a szomatikus mutációk meghatározása általában egy biopszia mintából történik. A daganatok intra-tumor heterogenitása miatt a szekvenálások során meghatározott mutációk egy jelentős része a vizsgált tumor régióra specifikusak. A jelenség szolid tumorokban már jól ismert, viszont az utóbbi években kiderült, hogy nem-szolid eredetű tumoroknál is jelen van [170]. Ráadásul, a tumor heterogenitás mértéke hatással volt a betegek túlélésére is [171].

In vitro inváziós kísérletekből végzett újgenerációs szekvenálás segítségével sikerült detektálni a mozgásból származó sejtösszetétel-változást. Az invázió mértéke 18,6% volt az A375 és a MEWO sejtvonal pároknál, míg az SK-MEL-28 és a MEWO sejtvonal pároknál csak 8,6% volt. Bár az SK-MEL-28 és az A375 sejtvonalak mozgása lényegesen nem tért el egymástól monokultúras körülmények között, az általuk elért invázió mértéke szignifikánsan eltért ($p=0,011$), amikor a homozigóta mutációk frekvenciáját hasonlítottam össze. Ezzel szemben a heterozigóta mutációkkal végzett elemzéseknél nem volt szignifikáns a különbség ($p=0,39$).

Kalibrációs szekvenálás során ismert összetételű sejtvonal párokat szekvenáltunk a módszer pontosságának meghatározására. A mutációk frekvenciájának a szórása egyre magasabb volt, ahogy a két sejtvonal aránya kiegyenlítődött. A legmagasabb szórás értéket két sejtvonal 50-50%-os arányánál kaptam. A heterozigóta mutációk szórásai lényegesen nagyobbak voltak, amely alacsony szekvenálási lefedettségénél tovább növekedett.

A mutációk frekvenciájának a szórása akár a 17%-ot is elérte mind a kalibrációs, mind az inváziós szekvenálásoknál. A kalibrációs szekvenálások technikai ismétlései között a szórások elérhették akár az 5,6%-ot. A szimulációk alapján a szórás 12,9% volt alacsonyabb (<50x) lefedettségénél. Nagyon magas 1000x szekvenálási lefedettségénél lényegesen kisebb a volt a szórás, ekkor 2% volt a legmagasabb érték.

A tumor sejtek motilitását vizsgáló kísérletek fő korlátja, hogy csak kétdimenziós sejtkultúrákkal lettek elvégezve. Az invázió során a sejteknek nem volt valódi ellenállása (pl. mátrix), a sejtek szinte szabadon mozoghattak, normál szöveti elemekkel nem

találkoztak. A tumorok növekedése akár több évig is tarthat, amíg eléri a klinikai detektálhatóság küszöbét. Ez idő alatt a sejtek lassan mozoghatnak, elvándorolhatnak. A mozgékonyabb, akár rosszabb prognózisú mutációval rendelkező sejtek így diszperz módon helyezkednek el a tumorban, nehezítve az azonosításukat.

A kísérletek további korlátozó tényezője, hogy egyszerre csak két sejtvonalat használtam. Ehhez képest egy betegben lévő tumorban több klón és akár több 100-1000 szub-klón található, melyek genetikai állománya bizonyos mértékben különbözik egymástól. További nehézségeket okozhat a tumorban jelen lévő normál szöveti kontamináció.

6.2. Az azonosított mutációk összefüggése a tumor minta méretével petefészek-tumoros betegekben

A vizsgálat célja a tumor minta mérete, valamint az azonosított mutációk száma közti kapcsolat meghatározása volt. Ehhez öt petefészek-tumoros betegből végeztünk multi-régió típusú szekvenálást, amely során a tumorból egy biopszia, biopszia körüli, és egész tumort átfogó mintákat gyűjtöttünk.

Két betegben azonosítottam öröklött mutációkat a BRCA1 illetve a BRCA2 génekben, amik hajlamosító tényezők a petefészek tumorok kialakulására. Ezeken kívül egy betegben az emlőrák kialakulását hajlamosító PALB2 génben azonosítottam mutációt [172, 173]. Mindegyik mutáció esetén szomatikusan törlődött a vad allél a tumorban.

A betegekben azonosított szomatikus mutációk leggyakrabban a TP53 és PTEN tumorszupresszor géneket érintették. Egy betegnél aktiváló mutációkat azonosítottam a PIK3CA és a KRAS onkogéneknél. Egy másik betegben a PI3K útvonal több pontos módosítással is aktiválódott, nála a PTEN gén inaktiváló mutációja, az FGFR2 gén aktiváló mutáció és két PIK3R1 gén inaktiváló mutációja volt azonosítható.

Megvizsgáltam a szomatikus mutációk mintázatát is, amely a mutációk kialakulásának mechanizmusáról adott információt. Mindegyik betegben jelen volt a korrall összefüggő 5-metilcitozin dezaminációjából származó mintázat. Egy betegben az APOBEC citidin-dezamináz hibával társuló magas C>T és C>G bázis cserével kapcsolatos mintázatot azonosítottam. Három betegben a hibás homológ rekombinációval társuló mintázat volt megfigyelhető.

Az egyes régiókban azonosított mutációk négy beteg esetén nagy mértékben megegyeztek. Esetükben a tumor minta méretének növelésével nem, vagy csak kevés többlet információt lehetett gyűjteni a tumorban lévő mutációkról. Egy betegben a kisebb tumor minta méretű biopsziában több mutáció volt azonosítható. Ennek hátterében feltehetően a tumor hipermutáló fenotípusa állt, mely következtében a sok felhalmozott mutáció kihígult, ahogy a tumor minta mérete növekedett. Érdekes megfigyelés volt, hogy a kópiaszám változás események szinte minden esetben megegyeztek, amiből arra lehet következtetni, hogy a kópiaszám változások a szomatikus mutációknál vagy konzervatívabb változások, vagy nehezebben azonosíthatók.

6.3. A Gentoype-2-Outcome (G-2-O) elemzőrendszer

Onkológiai vizsgálatok célja a betegek túlélésére, valamint kezelések hatékonyságára prognosztikus genetikai változások azonosítása. Emiatt gyakran sejt szabályozást, illetve fenotípust befolyásoló DNS, RNS vagy fehérje változást keresünk. Ezen megközelítés korlátja, hogy egy bizonyos fenotípus mögött többféle genetikai változás is lehet. Erre talán legjobb példa a mitogén-aktivált kináz útvonal aktivációja, ami történhet RTK amplifikáció/mutáció, KRAS vagy BRAF génmutációk következményeként.

Az általam fejlesztett Genotype-2-Outcome rendszer segítségével mutáció és génkifejeződés adatok együttes felhasználásával végeztem túlélés elemzést. A rendszerrel végzett túlélés elemzése során nem a mutációk direkt hatását vizsgálja, hanem a mutációkkal összefüggő génkifejeződési változások hatását. A GATA3 gén példáján látható volt, hogy a G-2-O által végzett túlélés elemzés emlőtumoros betegekben (HR=1,66, $p < 10^{-16}$) lényegesen szignifikánsabb, mint egyszerűen a mutációja ($p = 1,3 \times 10^{-8}$) vagy a kifejeződési szintje alapján (HR=0,71, $p = 1,3 \times 10^{-8}$) végzett túlélés elemzés.

Elemzőrendszerem egy hiánypótló módszer. Kevés olyan vizsgálat vagy adatbázis érhető el, mely esetén elegendően nagy számú tumoros beteg mutációiról és túléléséről kapunk együttesen információt. A TCGA adatbázisban lelhető fel a legtöbb emlőtumoros beteg adata, esetükben van információnk a gének kifejeződési szintjéről, a mutációkról és a kópiaszám változásokról is, viszont a betegek csak ~10%-ában érhető el túlélési adat.

A TP53 génmutációval rendelkező emlőtumoros betegekben több glikolízisben résztvevő gén (pl. PDK1, GP1, PGD) kifejeződési szintje emelkedett meg, viszont a

glikoneogenezisben részt vevő gének kifejeződése csökkent (pl. G6PC3, FBP1). A génkifejeződés változást RT-PCR-el segítségével validáltuk sejtvonalakban. Továbbá Seahorse XF96 extracellular flux analyzer eszköz felhasználásával bemutattuk, hogy a TP53 gén mutáns sejtvonalakban magasabb a glikolitikus és mitokondriális aktivitás a vad TP53 génnel rendelkező sejtekhez képest [154].

A tripla negatív emlő tumorok a legrosszabb prognózisú altípusba tartoznak. Ezen tumorokra célzott terápia még nem érhető el, a beteg egy része mégis jól reagál az antraciklin illetve a taxán alapú kezelésekre. Kollaborátoraink segítségével sikerült bemutatni, hogy azon betegek, akikben FOXA1/AR génmutációk azonosíthatók, jobb túléléssel rendelkeztek a TCGA adatbázis alapján ($p=0,05$), valamint jobban reagáltak antraciklin és taxán kezelésre ($p=0,02$) az MDACC (MD Anderson Cancer Center) adatok alapján. Ezen mutációk a BRCA hibajavító útvonalat érintették, emiatt a betegekben jóval magasabb volt a mutációs teher ($p<0,0001$). A BRCA deficienciához kötött génkifejeződési ujjlenyomat független adatsoron igazolva is jobb túléléssel társult ($p=0,009$) [149].

A TP53 gén mutációjakor sok sejtciklushoz kötött gén kifejeződési szintje emelkedett meg. Ezen gének közül 17 esetben találtam olyan publikációt, amely kifejezetten az adott gén kifejeződési szintjének elemzésével foglalkozott emlőtumoros adatokon. A gének magas kifejeződési szintje általában rossz prognózissal társult az összes emlőtumoros betegen külön, illetve kombináltan is (top 10 gén esetén: $HR=2,43$, $p<E-16$). Érdekes megfigyelés, hogy a kombinált mintázatuk hasonló jellegű volt, amikor a vad TP53 génre szűkítettem a betegeket ($HR=2,36$, $p=1,8E-3$), viszont megfordult a TP53 génmutációval rendelkező betegcsoportban ($HR=0,52$, $p=3,7E-2$).

Az elemzőrendszer hátránya, hogy kapcsolatot feltételez mutációk és génkifejeződés változás között. Bár sok esetben a feltételezés helyes, vannak példák, ahol bizonyos vezető génekben lévő mutációk nincsenek hatással a génkifejeződés szintjére, vagy az emelkedett gén kifejeződése nem növekei a fehérje mennyiségét.

7. Következtetések

Az utóbbi években több vizsgálat bemutatta, hogy magas a tumorok inter- és intra-tumor heterogenitása. Ez a genetikai diverzitás befolyásolja a klinikai szekvenálások eredményét, valamint nehezíti a prognosztikus és prediktív biomarkerek azonosítását.

Az eredményeim alapján levonható következtetések:

1. Újgenerációs szekvenálás felhasználásával követhető a mozgásból származó sejttösszetétel változás *in vitro* inváziós kísérletekben. Ismert összetételű sejtvonalak szekvenálása esetén pontosabban lehetett megállapítani az arányokat azon esetekben, amikor egy sejtvonal van túlsúlyban.
2. Mind az inváziós, mind a kalibrációs kísérletek során számított mutáció frekvenciák magas szórással rendelkeztek a biológiai ismétlések, valamint a technikai ismétlések között. A magas szórások jelentősen befolyásolhatják az elemzésekből levont következtetéseket, valamint további elemzéseket. Az *in silico* modellezésekből látható, hogy a szórás lényegesen csökken magas lefedettség esetén, így a klinikai diagnosztikai méréseket célszerű lehet magas lefedettségű szekvenálással végezni.
3. A nagyobb tumor részt lefedő szekvenálás nem befolyásolja a mutációk detektálását. Ez alól kivételek a hipermutáló tumorok, ahol kisebb tumor mintákban több szub-klonális mutáció azonosítható. Eredményeink azt tükrözik, hogy a jelenlegi klinikai diagnosztikai szekvenálások mintavételi protokolljai valószínűleg megbízható eredményt adnak, különösen, ha a szekvenálás mélysége elég nagy.
4. A szomatikus mutációk hatására gyakran nem csak egy fehérje, hanem az általa szabályozott jelátviteli útvonal, valamint a sejtben lévő RNS szintek is megváltoznak. A Genotype-2-Outcome elemzőrendszer felhasználásával bemutattam, hogy onkogén vagy tumorszupresszor gén mutációjához kötött génkifejeződés változások prognosztikusak a betegek túlélésére. A génkifejeződési változásokkal végzett túlélés elemzés sok esetben szignifikánsan jobb eredményt produkált az adott gén mutációjára, vagy kifejeződési szintje alapján elvégzett elemzéshez képest. Ezért ez a módszer új biomarker

azonosítására jó kiegészítést nyújthat a csak génmutáció vagy génkifejeződést vizsgáló rendszerek mellett.

5. A TP53 génmutációval rendelkező emlőtumoros betegekben több olyan gén kifejeződési szintje emelkedik meg, melyek a sejtosztódást szabályozzák. Ezen génekről több vizsgálat (tanulmány) bemutatta, hogy a magas kifejeződés rossz prognózissal jár. Vizsgálataim viszont azt valószínűsítik, hogy ezen gének magas kifejeződése a TP53 gén vad típusú betegekben rossz prognózissal, viszont a TP53 gén mutáns betegekben jobb a prognózissal jár. A megfigyelésekből következtethetünk arra, hogy ezen gének kifejeződése és a betegek prognózisának kapcsolata inkább indirekt módon tükrözi a TP53 gén mutációjának állapotát.

8. Összefoglalás

A tumorok genetikai heterogenitását több tényező befolyásolja, mint a növekedés során felhalmozott genetikai változások, a sejtek térbeli elhelyezkedése a tumorban, valamint a környezettel való interakció. Diagnosztikai szekvenálások során gyakran a tumor egy kis részletéből izolálnak DNS-t, melyből történik a szomatikus mutációk/változások meghatározása.

Munkám célja volt megvizsgálni, hogy újgenerációs szekvenálással milyen mértékben követhető a sejtvonalak mozgása *in vitro* invázió kísérletekben. Az A375 és az SK-MEL-28 sejtvonalak invázió mértéke lényegesen eltért egymástól ($p=0,011$). A sejtvonalak kalibrációs szekvenálása során láttuk, hogy egy sejten belül lévő homozigóta vagy heterozigóta mutációk frekvenciájának a szórása lényegesen nagyobb volt (8%-17% intervallum), mint a mutáció párok technikai ismétlései között (1,6%-5,6%). Az *in silico* modellezésekben látható, hogy a mutációk frekvenciájának a szórása csökkent magasabb szekvenálási lefedettségeknél. Az eredmények alapján kíváncsiabb magasabb lefedettséggel végezni a szekvenálást a pontosabb eredmény érdekében.

További célom volt meghatározni a szekvenált tumor minta méret és detektálható mutációk közötti összefüggést. Az eredményekben látható volt, hogy a méret növelésével lényegesen nem változott az azonosított mutációk mennyisége. Erre kivétel a hipermutációs fenotípussal rendelkező tumor volt, ahol kisebb tumor részlet szekvenálása során lényegesen több mutáció volt detektálható.

További célom volt egy olyan online elérhető rendszer készítése, amely egy gén mutációjához asszociált génkifejeződési mintázat alapján végez túlélés elemzést. Eredményeimből látható, hogy az emlőtumorkban leggyakrabban mutálódó gének kifejeződési ujjlenyomata erősen összefüggött a túléléssel (PIK3CA: $p=1,1E-06$, HR=1,6; TP53: $p<E-16$, HR=0,55; GATA3: $p<E-16$, HR=1,6; MAP3K1: $p<E-16$, HR=1,6; CDH1: $3,9E-9$, HR=1,4).

A TP53 génmutációjakor sok sejtciklust valamint DNS osztódást szabályozó gén kifejeződési szintje emelkedett meg. Ezen gének kifejeződési szintjük általában már önmagukban is prognosztikusak voltak, együttesen viszont lényegesen erősebb volt a hatásuk (HR=2,43, $p<E-16$).

9. Summary

Genetic heterogeneity of a tumor is affected by multiple factors, such as accumulated genetic alterations, spatial position of cells in the tumor, and interactions with the environment. This genetic diversity may affect clinical sequencing projects, where sequencing is usually performed using DNA extracted from a small section of the tumor.

The aim of my research was to examine the effects of cellular motility on next-generation sequencing using *in vitro* invasion assay experiments. Invasion of the A375 and the SK-MEL-28 cell lines differed significantly ($p=0,011$) in the experiments. By performing calibration sequencing, we observed that standard deviations of homozygous and heterozygous mutation frequencies were higher in biological replicates (8%-17%) compared to the technical replicates (1,6%-5,6%). Our *in silico* models have shown that increasing sequencing coverage decreased deviation of mutation frequencies. Based on these results, sequencing with high coverages can generate more reliable results.

In this work we used next generation sequencing to examine how isolated tumor sample size affects detectable genetic mutations. We found that increasing the size of the sequenced tumor size does not strongly affect detected mutations, with the exception of hypermutating tumors, where the number of identifiable genetic alterations decreased as sample size increased.

We developed an online tool (Genotype-2-Outcome) which is able to identify gene expression changes associated to a genetic alteration and perform survival analysis using the altered genes on an independent dataset in breast cancer patients. In our results we can see that gene expression alteration associated with the most commonly mutated genes in breast cancer is strongly associated with patient survival (PIK3CA: $p=1,1E-06$, HR=1,6; TP53: $p<E-16$, HR=0,55; GATA3: $p<E-16$, HR=1,6; MAP3K1: $p<E-16$, HR=1,6; CDH1: $3,9E-9$, HR=1,4).

Utilizing the Genotype-2-Outcome tool we found that expression of genes associated to cell cycle and DNA replication is elevated in breast cancer patients harboring mutations in the TP53 tumor suppressor gene. Expression of the identified genes are individually prognostic on the survival of patients, while their combination produced substantially better results (HR=2,43, $p<E-16$).

10. Irodalomjegyzék

1. Hanahan D, Weinberg RA. (2000) The hallmarks of cancer. *Cell*, 100: 57-70.
2. Hanahan D, Weinberg RA. (2011) Hallmarks of cancer: the next generation. *Cell*, 144: 646-674.
3. Hennessy BT, Smith DL, Ram PT, Lu Y, Mills GB. (2005) Exploiting the PI3K/AKT pathway for cancer drug discovery. *Nature reviews. Drug discovery*, 4: 988-1004.
4. Courtney KD, Corcoran RB, Engelman JA. (2010) The PI3K pathway as drug target in human cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 28: 1075-1083.
5. Brose MS, Volpe P, Feldman M, Kumar M, Rishi I, Gerrero R, Einhorn E, Herlyn M, Minna J, Nicholson A, Roth JA, Albelda SM, Davies H, Cox C, Brignell G, Stephens P, Futreal PA, Wooster R, Stratton MR, Weber BL. (2002) BRAF and RAS mutations in human lung cancer and melanoma. *Cancer research*, 62: 6997-7000.
6. Forrester K, Almoguera C, Han K, Grizzle WE, Perucho M. (1987) Detection of high incidence of K-ras oncogenes during human colon tumorigenesis. *Nature*, 327: 298-303.
7. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA, Leiserson MDM, Miller CA, Welch JS, Walter MJ, Wendl MC, Ley TJ, Wilson RK, Raphael BJ, Ding L. (2013) Mutational landscape and significance across 12 major cancer types. *Nature*, 502: 333-339.
8. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W, Bell R, Rosenthal J, Hussey C, Tran T, McClure M, Frye C, Hattier T, Phelps R, Haugen-Strano A, Katcher H, Yakumo K, Gholami Z, Shaffer D, Stone S, Bayer S, Wray C, Bogden R, Dayananth P, Ward J, Tonin P, Narod S, Bristow PK, Norris FH, Helvering L, Morrison P, Rosteck P, Lai M, Barrett JC, Lewis C, Neuhausen S, Cannon-Albright L, Goldgar D, Wiseman R, Kamb A, MH S. (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*, 266: 66-71.
9. Welch PL, King MC. (2001) BRCA1 and BRCA2 and the genetics of breast and ovarian cancer. *Human molecular genetics*, 10: 705-713.

10. Futreal PA, Liu Q, Shattuck-Eidens D, Cochran C, Harshman K, Tavtigian S, Bennett LM, Haugen-Strano A, Swensen J, Miki Y, Eddington K, McClure M, Frye C, Weaver-Feldhaus J, Ding W, Gholami Z, Soderkvist P, Terry L, Jhanwar S, Berchuck A, Iglehart JD, Marks J, Ballinger DG, Barrett JC, Skolnick MH, Kamb A, Wiseman R. (1994) BRCA1 mutations in primary breast and ovarian carcinomas. *Science*, 266: 120-122.
11. Tomlinson GE, Chen TT, Stastny VA, Virmani AK, Spillman MA, Tonk V, Blum JL, Schneider NR, Wistuba II, Shay JW, Minna JD, Gazdar AF. (1998) Characterization of a breast cancer cell line derived from a germ-line BRCA1 mutation carrier. *Cancer research*, 58: 3237-3242.
12. Esteller M, Silva JM, Dominguez G, Bonilla F, Matias-Guiu X, Lerma E, Bussaglia E, Prat J, Harkes IC, Repasky EA, Gabrielson E, Schutte M, Baylin SB, Herman JG. (2000) Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. *Journal of the National Cancer Institute*, 92: 564-569.
13. Holstege H, Joosse SA, van Oostrom CT, Nederlof PM, de Vries A, Jonkers J. (2009) High incidence of protein-truncating TP53 mutations in BRCA1-related breast cancer. *Cancer research*, 69: 3625-3633.
14. Rampino N, Yamamoto H, Ionov Y, Li Y, Sawai H, Reed JC, Perucho M. (1997) Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. *Science*, 275: 967-969.
15. Li J, Yen C, Liaw D, Podsypanina K, Bose S, Wang SI, Puc J, Miliarensis C, Rodgers L, McCombie R, Bigner SH, Giovanella BC, Ittmann M, Tycko B, Hibshoosh H, Wigler MH, Parsons R. (1997) PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science*, 275: 1943-1947.
16. Greenblatt MS, Bennett WP, Hollstein M, Harris CC. (1994) Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. *Cancer research*, 54: 4855-4878.
17. Bamford S, Dawson E, Forbes S, Clements J, Pettett R, Dogan A, Flanagan A, Teague J, Futreal PA, Stratton MR, Wooster R. (2004) The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *British journal of cancer*, 91: 355-358.
18. Nordling CO. (1953) A new theory on cancer-inducing mechanism. *British journal of cancer*, 7: 68-72.

19. Dunn JM, Phillips RA, Becker AJ, Gallie BL. (1988) Identification of germline and somatic mutations affecting the retinoblastoma gene. *Science*, 241: 1797-1800.
20. Miyoshi Y, Nagase H, Ando H, Horii A, Ichii S, Nakatsuru S, Aoki T, Miki Y, Mori T, Nakamura Y. (1992) Somatic mutations of the APC gene in colorectal tumors: mutation cluster region in the APC gene. *Human molecular genetics*, 1: 229-233.
21. Gnarr JR, Tory K, Weng Y, Schmidt L, Wei MH, Li H, Latif F, Liu S, Chen F, Duh FM, Lubensky I, Duan DR, Florence C, Pozzatti R, Walther MM, Bander NH, Grossman HB, Brauch H, Pomer S, Brooks JD, Isaacs WB, Lerman MI, Zbar B, Linehan WM. (1994) Mutations of the VHL tumour suppressor gene in renal carcinoma. *Nature genetics*, 7: 85-90.
22. Pelletier J, Bruening W, Li FP, Haber DA, Glaser T, Housman DE. (1991) WT1 mutations contribute to abnormal genital system development and hereditary Wilms' tumour. *Nature*, 353: 431-434.
23. Peto R, Darby S, Deo H, Silcocks P, Whitley E, Doll R. (2000) Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies. *BMJ*, 321: 323-329.
24. Pfeifer GP. (2010) Environmental exposures and mutational patterns of cancer genomes. *Genome medicine*, 2: 54.
25. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. (2013) Deciphering signatures of mutational processes operative in human cancer. *Cell reports*, 3: 246-259.
26. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL, Boyault S, Burkhardt B, Butler AP, Caldas C, Davies HR, Desmedt C, Eils R, Eyfjord JE, Foekens JA, Greaves M, Hosoda F, Hutter B, Ilcic T, Imbeaud S, Imielinski M, Jager N, Jones DT, Jones D, Knappskog S, Kool M, Lakhani SR, Lopez-Otin C, Martin S, Munshi NC, Nakamura H, Northcott PA, Pajic M, Papaemmanuil E, Paradiso A, Pearson JV, Puente XS, Raine K, Ramakrishna M, Richardson AL, Richter J, Rosenstiel P, Schlesner M, Schumacher TN, Span PN, Teague JW, Totoki Y, Tutt AN, Valdes-Mas R, van Buuren MM, van 't Veer L, Vincent-Salomon A, Waddell N, Yates LR, Zucman-Rossi J, Futreal PA, McDermott U, Lichter P, Meyerson M, Grimmond SM, Siebert R, Campo E, Shibata T, Pfister SM, Campbell PJ, Stratton MR. (2013) Signatures of mutational processes in human cancer. *Nature*, 500: 415-421.

27. Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, Nakamura Y, White R, Smits AM, Bos JL. (1988) Genetic alterations during colorectal-tumor development. *The New England journal of medicine*, 319: 525-532.
28. Fearon ER, Vogelstein B. (1990) A genetic model for colorectal tumorigenesis. *Cell*, 61: 759-767.
29. Morin PJ, Sparks AB, Korinek V, Barker N, Clevers H, Vogelstein B, Kinzler KW. (1997) Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. *Science*, 275: 1787-1790.
30. Samuels Y, Wang Z, Bardelli A, Silliman N, Ptak J, Szabo S, Yan H, Gazdar A, Powell SM, Riggins GJ, Willson JK, Markowitz S, Kinzler KW, Vogelstein B, Velculescu VE. (2004) High frequency of mutations of the PIK3CA gene in human cancers. *Science*, 304: 554.
31. Cantley LC, Neel BG. (1999) New insights into tumor suppression: PTEN suppresses tumor formation by restraining the phosphoinositide 3-kinase/AKT pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 96: 4240-4245.
32. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, Garnett MJ, Bottomley W, Davis N, Dicks E, Ewing R, Floyd Y, Gray K, Hall S, Hawes R, Hughes J, Kosmidou V, Menzies A, Mould C, Parker A, Stevens C, Watt S, Hooper S, Wilson R, Jayatilake H, Gusterson BA, Cooper C, Shipley J, Hargrave D, Pritchard-Jones K, Maitland N, Chenevix-Trench G, Riggins GJ, Bigner DD, Palmieri G, Cossu A, Flanagan A, Nicholson A, Ho JW, Leung SY, Yuen ST, Weber BL, Seigler HF, Darrow TL, Paterson H, Marais R, Marshall CJ, Wooster R, Stratton MR, Futreal PA. (2002) Mutations of the BRAF gene in human cancer. *Nature*, 417: 949-954.
33. Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, Cibulskis K, Sougnez C, Greulich H, Muzny DM, Morgan MB, Fulton L, Fulton RS, Zhang Q, Wendl MC, Lawrence MS, Larson DE, Chen K, Dooling DJ, Sabo A, Hawes AC, Shen H, Jhangiani SN, Lewis LR, Hall O, Zhu Y, Mathew T, Ren Y, Yao J, Scherer SE, Clerc K, Metcalf GA, Ng B, Milosavljevic A, Gonzalez-Garay ML, Osborne JR, Meyer R, Shi X, Tang Y, Koboldt DC, Lin L, Abbott R, Miner TL, Pohl C, Fewell G, Haipek C, Schmidt H, Dunford-Shore BH, Kraja A, Crosby SD, Sawyer CS, Vickery T, Sander S, Robinson J, Winckler W, Baldwin J, Chirieac LR, Dutt A,

Fennell T, Hanna M, Johnson BE, Onofrio RC, Thomas RK, Tonon G, Weir BA, Zhao X, Ziaugra L, Zody MC, Giordano T, Orringer MB, Roth JA, Spitz MR, Wistuba, II, Ozenberger B, Good PJ, Chang AC, Beer DG, Watson MA, Ladanyi M, Broderick S, Yoshizawa A, Travis WD, Pao W, Province MA, Weinstock GM, Varmus HE, Gabriel SB, Lander ES, Gibbs RA, Meyerson M, Wilson RK. (2008) Somatic mutations affect key pathways in lung adenocarcinoma. *Nature*, 455: 1069-1075.

34. Bell D., Berchuck A., Birrer M., Chien J., Cramer D., Dao F., Dhir R., DiSaia P., Gabra H., Glenn P., Godwin A., Gross J., Hartmann L., Huang M., Huntsman D., Iacocca M., Imielinski M., Kalloger S., Karlan B., Levine D., Mills G., Morrison C., Mutch D., Olvera N., Orsulic S., Park K., Petrelli N., Rabeno B., Rader J., Sikic B., Smith-McCune K., Sood A., Bowtell D., Penny R., Testa J., Chang K., Dinh H., Drummond J., Fowler G., Gunaratne P., Hawes A., Kovar C., Lewis L., Morgan M., Newsham I., Santibanez J., Reid J., Trevino L., Wu Y., Wang M., Muzny D., Wheeler D., Gibbs R., Getz G., Lawrence M., Cibulskis K., Sivachenko A., Sougnez C., Voet D., Wilkinson J., Bloom T., Ardlie K., Fennell T., Baldwin J., Gabriel S., Lander E., Ding LL., Fulton R., Koboldt D., McLellan M., Wylie T., Walker J., O'Laughlin M., Dooling D., Fulton L., Abbott R., Dees N., Zhang Q., Kandoth C., Wendl M., Schierding W., Shen D., Harris C., Schmidt H., Kalicki J., Delehaunty K., Fronick C., Demeter R., Cook L., Wallis J., Lin L., Magrini V., Hodges J., Eldred J., Smith S., Pohl C., Vandin F., Raphael B., Weinstock G., Mardis E., Wilson R., Meyerson M., Winckler W., Getz G., Verhaak R., Carter S., Mermel C., Saksena G., Nguyen H., Onofrio R., Lawrence M., Hubbard D., Gupta S., Crenshaw A., Ramos A., Ardlie K., Chin L., Protopopov A., Zhang J., Kim T., Perna I., Xiao Y., Zhang H., Ren G., Sathiamoorthy N., Park R., Lee E., Park P., Kucherlapati R., Absher M., Waite L., Sherlock G., Brooks J., Li J., Xu J., Myers R., Laird PW., Cope L., Herman J., Shen H., Weisenberger D., Noushmehr H., Pan F., Triche T Jr., Berman B., Van Den Berg D., Buckley J., Baylin S., Spellman P., Purdom E., Neuvial P., Bengtsson H., Jakkula L., Durinck S., Han J., Dorton S., Marr H., Choi Y., Wang V., Wang N., Ngai J., Conboy J., Parvin B., Feiler H., Speed T., Gray J., Levine A., Socci N., Liang Y., Taylor B., Schultz N., Borsu L., Lash A., Brennan C., Viale A., Sander C., Ladanyi M., Hoadley K., Meng S., Du Y., Shi Y., Li L., Turman Y., Zang D., Helms E., Balu S., Zhou X., Wu J., Topal M., Hayes D., Perou C., Getz G., Voet D., Saksena G., Zhang J., Zhang H., Wu C.,

- Shukla S., Cibulskis K., Lawrence M., Sivachenko A., Jing R., Park R., Liu Y., Park P., Noble M., Chin L., Carter H., Kim D., Karchin R., Spellman P., Purdom E., Neuvial P., Bengtsson H., Durinck S., Han J., Korkola J., Heiser L., Cho R., Hu Z., Parvin B., Speed T., Gray J., Schultz N., Cerami E., Taylor B., Olshen A., Reva B., Antipin Y., Shen R., Mankoo P., Sheridan R., Ciriello G., Chang W., Bernanke J., Borsu L., Levine D., Ladanyi M., Sander C., Haussler D., Benz C., Stuart J., Benz S., Sanborn J., Vaske C., Zhu J., Szeto C., Scott G., Yau C., Hoadley K., Du Y., Balu S., Hayes D., Perou C., Wilkerson M., Zhang N., Akbani R., Baggerly K., Yung W., Mills G., Weinstein J., Penny R., Shelton T., Grimm D., Hatfield M., Morris S., Yena P., Rhodes P., Sherman M., Paulauskis J., Millis S., Kahn A., Greene J., Sfeir R., Jensen M., Chen J., Whitmore J., Alonso S., Jordan J., Chu A., Zhang J., Barker A., Compton C., Eley G., Ferguson M., Fielding P., Gerhard D., Myles R., Schaefer C., Mills Shaw K., Vaught J., Vockley J., Good P., Guyer M., Ozenberger B., Peterson J., E. T. (2011) Integrated genomic analyses of ovarian carcinoma. *Nature*, 474: 609-615.
35. Armaghany T, Wilson JD, Chu Q, Mills G. (2012) Genetic alterations in colorectal cancer. *Gastrointestinal cancer research : GCR*, 5: 19-27.
 36. Sameer AS. (2013) Colorectal cancer: molecular mutations and polymorphisms. *Frontiers in oncology*, 3: 114.
 37. Muzny D, Bainbridge M, Chang K, Dinh H, Drummond J, Fowler G, Kovar C, Lewis L, Morgan M, Newsham I, Reid J, Santibanez J, Shinbrot E, Trevino L, Wu Y, Wang M, Gunaratne P, Donehower L, Creighton C, Wheeler D, Gibbs R, Lawrence M, Voet D, Jing R, Cibulskis K, Sivachenko A, Stojanov P, McKenna A, Lander E, Gabriel S, Getz G, Ding L, Fulton R, Koboldt D, Wylie T, Walker J, Dooling D, Fulton L, Delehaunty K, Fronick C, Demeter R, Mardis E, Wilson R, Chu A, Chun H, Mungall A, Pleasance E, Robertson A, Stoll D, Balasundaram M, Birol I, Butterfield Y, Chuah E, Coope R, Dhalla N, Guin R, Hirst C, Hirst M, Holt R, Lee D, Li H, Mayo M, Moore R, Schein J, Slobodan J, Tam A, Thiessen N, Varhol R, Zeng T, Zhao Y, Jones S, Marra M, Bass A, Ramos A, Saksena G, Cherniack A, Schumacher S, Tabak B, Carter S, Pho N, Nguyen H, Onofrio R, Crenshaw A, Ardlie K, Beroukhim R, Winckler W, Getz G, Meyerson M, Protopopov A, Zhang J, Hadjipanayis A, Lee E, Xi R, Yang L, Ren X, Zhang H, Sathiamoorthy N, Shukla S, Chen P, Haseley P, Xiao Y, Lee S, Seidman J, Chin L, Park P, Kucherlapati R, Auman J, Hoadley K, Du Y, Wilkerson M, Shi Y, Liquori

C, Meng S, Li L, Turman Y, Topal M, Tan D, Waring S, Buda E, Walsh J, Jones C, Mieczkowski P, Singh D, Wu J, Gulabani A, Dolina P, Bodenheimer T, Hoyle A, Simons J, Soloway M, Mose L, Jefferys S, Balu S, O'Connor B, Prins J, Chiang D, Hayes D, Perou C, Hinoue T, Weisenberger D, Maglinte D, Pan F, Berman B, Van D, Berg, DJ., Shen H, Triche T, Jr., Baylin S, Laird P, Getz G, Noble M, Voet D, Saksena G, Gehlenborg N, DiCara D, Zhang J, Zhang H, Wu C, Liu S, Shukla S, Lawrence M, Zhou L, Sivachenko A, Lin P, Stojanov P, Jing R, Park R, Nazaire M, Robinson J, Thorvaldsdottir H, Mesirov J, Park P, Chin L, Thorsson V, Reynolds S, Bernard B, Kreisberg R, Lin J, Iype L, Bressler R, Erkkilä T, Gundapuneni M, Liu Y, Norberg A, Robinson T, Yang D, Zhang W, Shmulevich I, de R, JJ., Schultz N, Cerami E, Ciriello G, Goldberg A, Gross B, Jacobsen A, Gao J, Kaczkowski B, Sinha R, Aksoy B, Antipin Y, Reva B, Shen R, Taylor B, Chan T, Ladanyi M, Sander C, Akbani R, Zhang N, Broom B, Casasent T, Unruh A, Wakefield C, Hamilton S, Cason R, Baggerly K, Weinstein J, Haussler D, Benz C, Stuart J, Benz S, Sanborn J, Vaske C, Zhu J, Szeto C, Scott G, Yau C, Ng S, Goldstein T, Ellrott K, Collisson E, Cozen A, Zerbino D, Wilks C, Craft B, Spellman P, Penny R, Shelton T, Hatfield M, Morris S, Yena P, Shelton C, Sherman M, Paulauskis J, Gastier-Foster J, Bowen J, Ramirez N, Black A, Pyatt R, Wise L, White P, Bertagnolli M, Brown J, Chan T, Chu G, Czerwinski C, Denstman F, Dhir R, Dörner A, Fuchs C, Guillem J, Iacocca M, Juhl H, Kaufman A, Kohl B, 3rd., Van L, X., Mariano M, Medina E, Meyers M, Nash G, Paty P, Petrelli N, Rabeno B, Richards W, Solit D, Swanson P, Temple L, Tepper J, Thorp R, Vakiani E, Weiser M, Willis J, Witkin G, Zeng Z, Zinner M, Zornig C, Jensen M, Sfeir R, Kahn A, Chu A, Kothiyal P, Wang Z, Snyder E, Pontius J, Pihl T, Ayala B, Backus M, Walton J, Whitmore J, Baboud J, Berton D, Nicholls M, Srinivasan D, Raman R, Girshik S, Kigonya P, Alonso S, Sanbhadti R, Barletta S, Greene J, Pot D, Shaw K, Dillon L, Buetow K, Davidsen T, Demchok J, Eley G, Ferguson M, Fielding P, Schaefer C, Sheth M, Yang L, Guyer M, Ozenberger B, Palchik J, Peterson J, Sofia H, Thomson E. (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, 487: 330-337.

38. Easton DF, Ford D, Bishop DT. (1995) Breast and ovarian cancer incidence in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. *American journal of human genetics*, 56: 265-271.

39. Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL, Loman N, Olsson H, Johannsson O, Borg A, Pasini B, Radice P, Manoukian S, Eccles DM, Tang N, Olah E, Anton-Culver H, Warner E, Lubinski J, Gronwald J, Gorski B, Tulinius H, Thorlacius S, Eerola H, Nevanlinna H, Syrjakoski K, Kallioniemi OP, Thompson D, Evans C, Peto J, Lalloo F, Evans DG, Easton DF. (2003) Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. *American journal of human genetics*, 72: 1117-1130.
40. Antoniou AC, Casadei S, Heikkinen T, Barrowdale D, Pylkas K, Roberts J, Lee A, Subramanian D, De Leeneer K, Fostira F, Tomiak E, Neuhausen SL, Teo ZL, Khan S, Aittomaki K, Moilanen JS, Turnbull C, Seal S, Mannermaa A, Kallioniemi A, Lindeman GJ, Buys SS, Andrulis IL, Radice P, Tondini C, Manoukian S, Toland AE, Miron P, Weitzel JN, Domchek SM, Poppe B, Claes KB, Yannoukakos D, Concannon P, Bernstein JL, James PA, Easton DF, Goldgar DE, Hopper JL, Rahman N, Peterlongo P, Nevanlinna H, King MC, Couch FJ, Southey MC, Winqvist R, Foulkes WD, Tischkowitz M. (2014) Breast-cancer risk in families with mutations in PALB2. *The New England journal of medicine*, 371: 497-506.
41. Moynahan ME, Chiu JW, Koller BH, Jasin M. (1999) Brca1 controls homology-directed DNA repair. *Molecular cell*, 4: 511-518.
42. Scully R, Chen J, Plug A, Xiao Y, Weaver D, Feunteun J, Ashley T, Livingston DM. (1997) Association of BRCA1 with Rad51 in mitotic and meiotic cells. *Cell*, 88: 265-275.
43. Scully R, Livingston DM. (2000) In search of the tumour-suppressor functions of BRCA1 and BRCA2. *Nature*, 408: 429-432.
44. Osborne CK, Yochmowitz MG, Knight WA, 3rd, McGuire WL. (1980) The value of estrogen and progesterone receptors in the treatment of breast cancer. *Cancer*, 46: 2884-2888.
45. Clarke M, Collins R, Darby S, Davies C, Evans V, Godwin J, Gray R, McGale P, Peto R, Wang Y. (2005) Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet*, 365: 1687-1717.
46. Onitilo AA, Engel JM, Greenlee RT, Mukesh BN. (2009) Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival. *Clinical medicine & research*, 7: 4-13.

47. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, Karaca G, Troester MA, Tse CK, Edmiston S, Deming SL, Geradts J, Cheang MC, Nielsen TO, Moorman PG, Earp HS, Millikan RC. (2006) Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA*, 295: 2492-2502.
48. Prat A, Perou CM. (2011) Deconstructing the molecular portraits of breast cancer. *Molecular oncology*, 5: 5-23.
49. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ. (2011) Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Annals of oncology : official journal of the European Society for Medical Oncology*, 22: 1736-1747.
50. Romond EH, Perez EA, Bryant J, Suman VJ, Geyer CE, Jr., Davidson NE, Tan-Chiu E, Martino S, Paik S, Kaufman PA, Swain SM, Pisansky TM, Fehrenbacher L, Kutteh LA, Vogel VG, Visscher DW, Yothers G, Jenkins RB, Brown AM, Dakhil SR, Mamounas EP, Lingle WL, Klein PM, Ingle JN, Wolmark N. (2005) Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *The New England journal of medicine*, 353: 1673-1684.
51. Piccart-Gebhart MJ, Procter M, Leyland-Jones B, Goldhirsch A, Untch M, Smith I, Gianni L, Baselga J, Bell R, Jackisch C, Cameron D, Dowsett M, Barrios CH, Steger G, Huang CS, Andersson M, Inbar M, Lichinitser M, Lang I, Nitz U, Iwata H, Thomssen C, Lohrisch C, Suter TM, Ruschoff J, Suto T, Giatromanolaki S, Ward C, Straehle C, McFadden E, Dolci MS, Gelber RD. (2005) Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *The New England journal of medicine*, 353: 1659-1672.
52. Finn RS, Crown JP, Lang I, Boer K, Bondarenko IM, Kulyk SO, Ettl J, Patel R, Pinter T, Schmidt M, Shparyk Y, Thummala AR, Voytko NL, Fowst C, Huang X, Kim ST, Randolph S, Slamon DJ. (2015) The cyclin-dependent kinase 4/6 inhibitor palbociclib in combination with letrozole versus letrozole alone as first-line treatment of oestrogen receptor-positive, HER2-negative, advanced breast cancer (PALOMA-1/TRIO-18): a randomised phase 2 study. *The Lancet. Oncology*, 16: 25-35.
53. Fisher B, Costantino J, Redmond C, Poisson R, Bowman D, Couture J, Dimitrov NV, Wolmark N, Wickerham DL, Fisher ER, Margolese R, Robidoux A, Shibata H, Terz J, Paterson A, Feldman MI, Farrar W, Evans J, Lickley HL, Ketner M.

- (1989) A randomized clinical trial evaluating tamoxifen in the treatment of patients with node-negative breast cancer who have estrogen-receptor-positive tumors. *The New England journal of medicine*, 320: 479-484.
54. Ellis MJ, Coop A, Singh B, Mauriac L, Llombert-Cussac A, Janicke F, Miller WR, Evans DB, Dugan M, Brady C, Quebe-Fehling E, Borgs M. (2001) Letrozole is more effective neoadjuvant endocrine therapy than tamoxifen for ErbB-1- and/or ErbB-2-positive, estrogen receptor-positive primary breast cancer: evidence from a phase III randomized trial. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 19: 3808-3816.
 55. Kaufman B, Mackey JR, Clemens MR, Bapsy PP, Vaid A, Wardley A, Tjulandin S, Jahn M, Lehle M, Feyereislova A, Revil C, Jones A. (2009) Trastuzumab plus anastrozole versus anastrozole alone for the treatment of postmenopausal women with human epidermal growth factor receptor 2-positive, hormone receptor-positive metastatic breast cancer: results from the randomized phase III TAnDEM study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 27: 5529-5537.
 56. Goss PE, Ingle JN, Pritchard KI, Ellis MJ, Sledge GW, Budd GT, Rabaglio M, Ansari RH, Johnson DB, Tozer R, D'Souza DP, Chalchal H, Spadafora S, Stearns V, Perez EA, Liedke PE, Lang I, Elliott C, Gelmon KA, Chapman JA, Shepherd LE. (2013) Exemestane versus anastrozole in postmenopausal women with early breast cancer: NCIC CTG MA.27--a randomized controlled phase III trial. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 31: 1398-1404.
 57. Baselga J, Semiglazov V, van Dam P, Manikhas A, Bellet M, Mayordomo J, Campone M, Kubista E, Greil R, Bianchi G, Steinseifer J, Molloy B, Tokaji E, Gardner H, Phillips P, Stumm M, Lane HA, Dixon JM, Jonat W, Rugo HS. (2009) Phase II randomized study of neoadjuvant everolimus plus letrozole compared with placebo plus letrozole in patients with estrogen receptor-positive breast cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 27: 2630-2637.
 58. Koboldt DC., Fulton RS., McLellan MD., Schmidt H., Kalicki-Veizer J., McMichael JF., Fulton LL., Dooling DJ., Ding L., Mardis ER., Wilson RK., Ally A., Balasundaram M., Butterfield YS., Carlsen R., Carter C., Chu A., Chuah E., Chun HJ., Coope RJ., Dhalla N., Guin R., Hirst C., Hirst M., Holt RA., Lee D., Li

HI., Mayo M., Moore RA., Mungall AJ., Pleasance E., Robertson A., Schein JE.,
 Shafiei A., Sipahimalani P., Slobodan JR., Stoll D., Tam A., Thiessen N., Varhol
 RJ., Wye N., Zeng T., Zhao Y., Birol I., Jones SJ., Marra MA., Cherniack AD.,
 Saksena G., Onofrio RC., Pho NH., Carter SL., Schumacher SE., Tabak B.,
 Hernandez B., Gentry J., Nguyen H., Crenshaw A., Ardlie K., Beroukheim R.,
 Winckler W., Getz G., Gabriel SB., Meyerson M., Chin L., Park PJ., Kucherlapati
 R., Hoadley KA., Auman J., Fan C., Turman YJ., Shi Y., Li L., Topal MD., He X.,
 Chao HH., Prat A., Silva GO., Iglesia MD., Zhao W., Usary J., Berg JS., Adams
 M., Booker J., Wu J., Gulabani A., Bodenheimer T., Hoyle AP., Simons JV.,
 Soloway MG., Mose LE., Jefferys SR., Balu S., Parker JS., Hayes D., Perou CM.,
 Malik S., Mahurkar S., Shen H., Weisenberger DJ., Triche T Jr., Lai PH., Bootwalla
 MS., Maglinte DT., Berman BP., Van Den Berg DJ., Baylin SB., Laird PW.,
 Creighton CJ., Donehower LA., Getz G., Noble M., Voet D., Saksena G.,
 Gehlenborg N., DiCara D., Zhang J., Zhang H., Wu CJ., Liu SY., Lawrence MS.,
 Zou L., Sivachenko A., Lin P., Stojanov P., Jing R., Cho J., Sinha R., Park RW.,
 Nazaire MD., Robinson J., Thorvaldsdottir H., Mesirov J., Park PJ., Chin L.,
 Reynolds S., Kreisberg RB., Bernard B., Bressler R., Erkkila T., Lin J., Thorsson
 V., Zhang W., Shmulevich I., Ciriello G., Weinhold N., Schultz N., Gao J., Cerami
 E., Gross B., Jacobsen A., Sinha R., Aksoy B., Antipin Y., Reva B., Shen R., Taylor
 BS., Ladanyi M., Sander C., Anur P., Spellman PT., Lu Y., Liu W., Verhaak RR.,
 Mills GB., Akbani R., Zhang N., Broom BM., Casasent TD., Wakefield C., Unruh
 AK., Baggerly K., Coombes K., Weinstein JN., Haussler D., Benz CC., Stuart JM.,
 Benz SC., Zhu J., Szeto CC., Scott GK., Yau C., Paull EO., Carlin D., Wong C.,
 Sokolov A., Thusberg J., Mooney S., Ng S., Goldstein TC., Ellrott K., Grifford M.,
 Wilks C., Ma S., Craft B., Yan C., Hu Y., Meerzaman D., Gastier-Foster JM.,
 Bowen J., Ramirez NC., Black AD., Pyatt RE., White P., Zmuda EJ., Frick J.,
 Lichtenberg TM., Brookens R., George MM., Gerken MA., Harper HA., Leraas
 KM., Wise LJ., Tabler TR., McAllister C., Barr T., Hart-Kothari M., Tarvin K.,
 Saller C., Sandusky G., Mitchell C., Iacocca MV., Brown J., Rabeno B., Czerwinski
 C., Petrelli N., Dolzhansky O., Abramov M., Voronina O., Potapova O., Marks JR.,
 Suchorska WM., Murawa D., Kycler W., Ibbs M., Korski K., Spychała A., Murawa
 P., Brzeziński JJ., Perz H., Łażniak R., Teresiak M., Tatka H., Leporowska E.,
 Bogusz-Czerniewicz M., Malicki J., Mackiewicz A., Wiznerowicz M., Le XV.,
 Kohl B., Nguyen VT., Thorp R., Nguyen VB., Sussman H., Bui DP., Hajek R.,

- Nguyen PH., Tran VT., Huynh QT., Khan KZ., Penny R., Mallery D., Curley E., Shelton C., Yena P., Ingle JN., Couch FJ., Lingle WL., King TA., Gonzalez-Angulo AM., Mills GB., Dyer MD., Liu S., Meng X., Patangan M., Waldman F., Stöppler H., Rathmell W., Thorne L., Huang M., Boice L., Hill A., Morrison C., Gaudioso C., Bshara W., Daily K., Egea SC., Pegram M., Gomez-Fernandez C., Dhir R., Bhargava R., Brufsky A., Shriver CD., Hooke JA., Campbell JL., Mural RJ., Hu H., Somiari S., Larson C., Deyarmin B., Kvecher L., Kovatich AJ., Ellis MJ., King TA., Hu H., Couch FJ., Mural RJ., Stricker T., White K., Olopade O., Ingle JN., Luo C., Chen Y., Marks JR., Waldman F., Wiznerowicz M., Bose R., Chang LW., Beck AH., Gonzalez-Angulo AM., Pihl T., Jensen M., Sfeir R., Kahn A., Chu A., Kothiyal P., Wang Z., Snyder E., Pontius J., Ayala B., Backus M., Walton J., Baboud J., Berton D., Nicholls M., Srinivasan D., Raman R., Girshik S., Kigonya P., Alonso S., Sanbhadti R., Barletta S., Pot D., Sheth M., Demchok JA., Shaw KR., Yang L., Eley G., Ferguson ML., Tarnuzzer RW., Zhang J., Dillon LA., Buetow K., Fielding P., Ozenberger BA., Guyer MS., Sofia HJ., JD. P. (2012) Comprehensive molecular portraits of human breast tumours. *Nature*, 490: 61-70.
59. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, Martincorena I, Alexandrov LB, Martin S, Wedge DC, Van Loo P, Ju YS, Smid M, Brinkman AB, Morganella S, Aure MR, Lingjaerde OC, Langerod A, Ringner M, Ahn SM, Boyault S, Brock JE, Broeks A, Butler A, Desmedt C, Dirix L, Dronov S, Fatima A, Foekens JA, Gerstung M, Hooijer GK, Jang SJ, Jones DR, Kim HY, King TA, Krishnamurthy S, Lee HJ, Lee JY, Li Y, McLaren S, Menzies A, Mustonen V, O'Meara S, Pauporte I, Pivot X, Purdie CA, Raine K, Ramakrishnan K, Rodriguez-Gonzalez FG, Romieu G, Sieuwerts AM, Simpson PT, Shepherd R, Stebbings L, Stefansson OA, Teague J, Tommasi S, Treilleux I, Van den Eynden GG, Vermeulen P, Vincent-Salomon A, Yates L, Caldas C, van't Veer L, Tutt A, Knappskog S, Tan BK, Jonkers J, Borg A, Ueno NT, Sotiriou C, Viari A, Futreal PA, Campbell PJ, Span PN, Van Laere S, Lakhani SR, Eyfjord JE, Thompson AM, Birney E, Stunnenberg HG, van de Vijver MJ, Martens JW, Borresen-Dale AL, Richardson AL, Kong G, Thomas G, Stratton MR. (2016) Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*, 534: 47-54.
60. Piepkorn M. (2000) Melanoma genetics: an update with focus on the CDKN2A(p16)/ARF tumor suppressors. *Journal of the American Academy of Dermatology*, 42: 705-722; quiz 723-706.

61. Palmer JS, Duffy DL, Box NF, Aitken JF, O'Gorman LE, Green AC, Hayward NK, Martin NG, Sturm RA. (2000) Melanocortin-1 receptor polymorphisms and risk of melanoma: is the association explained solely by pigmentation phenotype? *American journal of human genetics*, 66: 176-186.
62. Garbe C, Buttner P, Weill J, Soyer HP, Stocker U, Kruger S, Roser M, Weckbecker J, Panizzon R, Bahmer F, Tilgen W, Guggenmoos-Holzmam I, Orfanos CE. (1994) Risk factors for developing cutaneous melanoma and criteria for identifying persons at risk: multicenter case-control study of the Central Malignant Melanoma Registry of the German Dermatological Society. *The Journal of investigative dermatology*, 102: 695-699.
63. Akbani R, Akdemir K, Aksoy B, Albert M, Ally A, Amin S, Arachchi H, Arora A, Auman J, Ayala B, Baboud J, Balasundaram M, Balu S, Barnabas N, Bartlett J, Bartlett P, Bastian B, Baylin S, Behera M, Belyaev D, Benz C, Bernard B, Beroukhir R, Bir N, Black A, Bodenheimer T, Boice L, Boland G, Bono R, Bootwalla M, Bosenberg M, Bowen J, Bowlby R, Bristow C, Brockway-Lunardi L, Brooks D, Brzezinski J, Bshara W, Buda E, Burns W, Butterfield Y, Button M, Calderone T, Cappellini G, Carter C, Carter S, Cherney L, Cherniack A, Chevalier A, Chin L, Cho J, Cho R, Choi Y, Chu A, Chudamani S, Cibulskis K, Ciriello G, Clarke A, Coons S, Cope L, Crain D, Curley E, Danilova L, D'Atri S, Davidsen T, Davies M, Delman K, Demchok J, Deng Q, Deribe Y, Dhalla N, Dhir R, DiCara D, Dinikin M, Dubina M, Ebrom J, Egea S, Eley G, Engel J, Eschbacher J, Fedosenko K, Felau I, Fennell T, Ferguson M, Fisher S, Flaherty K, Frazer S, Frick J, Fulidou V, Gabriel S, Gao J, Gardner J, Garraway L, Gastier-Foster J, Gaudioso C, Gehlenborg N, Genovese G, Gerken M, Gershenwald J, Getz G, Gomez-Fernandez C, Gribbin T, Grimsby J, Gross B, Guin R, Gutschner T, Hadjipanayis A, Halaban R, Hanf B, Haussler D, Haydu L, Hayes D, Hayward N, Heiman D, Herbert L, Herman J, Hersey P, Hoadley K, Hodis E, Holt R, Hoon D, Hoppough S, Hoyle A, Huang F, Huang M, Huang S, Hutter C, Ibbs M, Iype L, Jacobsen A, Jakrot V, Janning A, Jeck W, Jefferys S, Jensen M, Jones C, Jones S, Ju Z, Kakavand H, Kang H, Kefford R, Khuri F, Kim J, Kirkwood J, Klode J, Korkut A, Korski K, Krauthammer M, Kucherlapati R, Kwong L, Kyceler W, Ladanyi M, Lai P, Laird P, Lander E, Lawrence M, Lazar A, Łażniak R, Lee D, Lee J, Lee J, Lee K, Lee S, Lee W, Leporowska E, Leraas K, Li H, Lichtenberg T, Lichtenstein L, Lin P, Ling S, Liu J, Liu O, Liu W, Long G, Lu Y, Ma S, Ma Y, Mackiewicz A, Mahadeshwar H,

Malke J, Mallery D, Manikhas G, Mann G, Marra M, Matejka B, Mayo M, Mehrabi S, Meng S, Meyerson M, Mieczkowski P, Miller J, Miller M, Mills G, Moiseenko F, Moore R, Morris S, Morrison C, Morton D, Moschos S, Mose L, Muller F, Mungall A, Murawa D, Murawa P, Murray B, Nezi L, Ng S, Nicholson D, Noble M, Osunkoya A, Owonikoko T, Ozenberger B, Pagani E, Paklina O, Pantazi A, Parfenov M, Parfitt J, Park P, Park W, Parker J, Passarelli F, Penny R, Perou C, Pihl T, Potapova O, Prieto V, Protopopov A, Quinn M, Radenbaugh A, Rai K, Ramalingam S, Raman A, Ramirez N, Ramirez R, Rao U, Rathmell W, Ren X, Reynolds S, Roach J, Robertson A, Ross M, Roszik J, Russo G, Saksena G, Saller C, Samuels Y, Sander C, Sander C, Sandusky G, Santoso N, Saul M, Saw R, Schadendorf D, Schein J, Schultz N, Schumacher S, Schwallier C, Scolyer R, Seidman J, Sekhar P, Sekhon H, Senbabaoglu Y, Seth S, Shannon K, Sharpe S, Sharpless N, Shaw K, Shelton C, Shelton T, Shen R, Sheth M, Shi Y, Shiau C, Shmulevich I, Sica G, Simons J, Sinha R, Sipahimalani P, Sofia H, Soloway M, Song X, Sougnez C, Spillane A, Spychala A, Stretch J, Stuart J, Suchorska W, Sucker A, Sumer S, Sun Y, Synott M, Tabak B, Tabler T, Tam A, Tan D, Tang J, Tarnuzzer R, Tarvin K, Tatka H, Taylor B, Teresiak M, Thiessen N, Thompson J, Thorne L, Thorsson V, Trent J, Triche T, Jr., Tsai K, Tsou P, Van D, Berg, DJ., Van A, EM., Veluvolu U, Verhaak R, Voet D, Voronina O, Walter V, Walton J, Wan Y, Wang Y, Wang Z, Waring S, Watson I, Weinhold N, Weinstein J, Weisenberger D, White P, Wilkerson M, Wilmott J, Wise L, Wiznerowicz M, Woodman S, Wu C, Wu C, Wu J, Wu Y, Xi R, Xu A, Yang D, Yang L, Yang L, Zack T, Zenklusen J, Zhang H, Zhang J, Zhang W, Zhao X, Zhu J, Zhu K, Zimmer L, Zmuda E, Zou L. (2015) Genomic Classification of Cutaneous Melanoma. *Cell*, 161: 1681-1696.

64. Chalmers ZR, Connelly CF, Fabrizio D, Gay L, Ali SM, Ennis R, Schrock A, Campbell B, Shlien A, Chmielecki J, Huang F, He Y, Sun J, Tabori U, Kennedy M, Lieber DS, Roels S, White J, Otto GA, Ross JS, Garraway L, Miller VA, Stephens PJ, Frampton GM. (2017) Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome medicine*, 9: 34.
65. Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, Dummer R, Garbe C, Testori A, Maio M, Hogg D, Lorigan P, Lebbe C, Jouary T, Schadendorf D, Ribas A, O'Day SJ, Sosman JA, Kirkwood JM, Eggermont AM, Dreno B, Nolop K, Li J, Nelson B, Hou J, Lee RJ, Flaherty KT, McArthur GA.

- (2011) Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *The New England journal of medicine*, 364: 2507-2516.
66. Hauschild A, Grob JJ, Demidov LV, Jouary T, Gutzmer R, Millward M, Rutkowski P, Blank CU, Miller WH, Jr., Kaempgen E, Martin-Algarra S, Karaszewska B, Mauch C, Chiarion-Sileni V, Martin AM, Swann S, Haney P, Mirakhur B, Guckert ME, Goodman V, Chapman PB. (2012) Dabrafenib in BRAF-mutated metastatic melanoma: a multicentre, open-label, phase 3 randomised controlled trial. *Lancet*, 380: 358-365.
 67. Robert C, Karaszewska B, Schachter J, Rutkowski P, Mackiewicz A, Stroiakovski D, Lichinitser M, Dummer R, Grange F, Mortier L, Chiarion-Sileni V, Drucis K, Krajsova I, Hauschild A, Lorigan P, Wolter P, Long GV, Flaherty K, Nathan P, Ribas A, Martin AM, Sun P, Crist W, Legos J, Rubin SD, Little SM, Schadendorf D. (2015) Improved overall survival in melanoma with combined dabrafenib and trametinib. *The New England journal of medicine*, 372: 30-39.
 68. Larkin J, Ascierto PA, Dreno B, Atkinson V, Liskay G, Maio M, Mandala M, Demidov L, Stroyakovskiy D, Thomas L, de la Cruz-Merino L, Dutriaux C, Garbe C, Sovak MA, Chang I, Choong N, Hack SP, McArthur GA, Ribas A. (2014) Combined vemurafenib and cobimetinib in BRAF-mutated melanoma. *The New England journal of medicine*, 371: 1867-1876.
 69. King MC, Marks JH, Mandell JB. (2003) Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. *Science*, 302: 643-646.
 70. Risch HA, McLaughlin JR, Cole DE, Rosen B, Bradley L, Kwan E, Jack E, Vesprini DJ, Kuperstein G, Abrahamson JL, Fan I, Wong B, Narod SA. (2001) Prevalence and penetrance of germline BRCA1 and BRCA2 mutations in a population series of 649 women with ovarian cancer. *American journal of human genetics*, 68: 700-710.
 71. Kanchi KL, Johnson KJ, Lu C, McLellan MD, Leiserson MD, Wendl MC, Zhang Q, Koboldt DC, Xie M, Kandoth C, McMichael JF, Wyczalkowski MA, Larson DE, Schmidt HK, Miller CA, Fulton RS, Spellman PT, Mardis ER, Druley TE, Graubert TA, Goodfellow PJ, Raphael BJ, Wilson RK, Ding L. (2014) Integrated analysis of germline and somatic variants in ovarian cancer. *Nature communications*, 5: 3156.
 72. Burger RA, Brady MF, Bookman MA, Fleming GF, Monk BJ, Huang H, Mannel RS, Homesley HD, Fowler J, Greer BE, Boente M, Birrer MJ, Liang SX. (2011)

Incorporation of bevacizumab in the primary treatment of ovarian cancer. *The New England journal of medicine*, 365: 2473-2483.

73. Audeh MW, Carmichael J, Penson RT, Friedlander M, Powell B, Bell-McGuinn KM, Scott C, Weitzel JN, Oaknin A, Loman N, Lu K, Schmutzler RK, Matulonis U, Wickens M, Tutt A. (2010) Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer: a proof-of-concept trial. *Lancet*, 376: 245-251.
74. Morgan RJ, Jr., Armstrong DK, Alvarez RD, Bakkum-Gamez JN, Behbakht K, Chen LM, Copeland L, Crispens MA, DeRosa M, Dorigo O, Gershenson DM, Gray HJ, Hakam A, Havrilesky LJ, Johnston C, Lele S, Martin L, Matulonis UA, O'Malley DM, Penson RT, Percac-Lima S, Pineda M, Plaxe SC, Powell MA, Ratner E, Remmenga SW, Rose PG, Sabbatini P, Santoso JT, Werner TL, Burns J, Hughes M. (2016) Ovarian Cancer, Version 1.2016, NCCN Clinical Practice Guidelines in Oncology. *Journal of the National Comprehensive Cancer Network : JNCCN*, 14: 1134-1163.
75. Gradishar WJ, Anderson BO, Balassanian R, Blair SL, Burstein HJ, Cyr A, Elias AD, Farrar WB, Forero A, Giordano SH, Goetz MP, Goldstein LJ, Isakoff SJ, Lyons J, Marcom PK, Mayer IA, McCormick B, Moran MS, O'Regan RM, Patel SA, Pierce LJ, Reed EC, Salerno KE, Schwartzberg LS, Sitapati A, Smith KL, Smith ML, Soliman H, Somlo G, Telli M, Ward JH, Shead DA, Kumar R. (2017) NCCN Guidelines Insights: Breast Cancer, Version 1.2017. *Journal of the National Comprehensive Cancer Network : JNCCN*, 15: 433-451.
76. Coit DG, Thompson JA, Algazi A, Andtbacka R, Bichakjian CK, Carson WE, 3rd, Daniels GA, DiMaio D, Ernstoff M, Fields RC, Fleming MD, Gonzalez R, Guild V, Halpern AC, Hodi FS, Jr., Joseph RW, Lange JR, Martini MC, Materin MA, Olszanski AJ, Ross MI, Salama AK, Skitzki J, Sosman J, Swetter SM, Tanabe KK, Torres-Roca JF, Trisal V, Urist MM, McMillian N, Engh A. (2016) Melanoma, Version 2.2016, NCCN Clinical Practice Guidelines in Oncology. *Journal of the National Comprehensive Cancer Network : JNCCN*, 14: 450-473.
77. Housman G, Byler S, Heerboth S, Lapinska K, Longacre M, Snyder N, Sarkar S. (2014) Drug resistance in cancer: an overview. *Cancers*, 6: 1769-1792.
78. Gazdar AF. (2009) Activating and resistance mutations of EGFR in non-small-cell lung cancer: role in clinical response to EGFR tyrosine kinase inhibitors. *Oncogene*, 28 Suppl 1: S24-31.

79. Stewart EL, Tan SZ, Liu G, Tsao MS. (2015) Known and putative mechanisms of resistance to EGFR targeted therapies in NSCLC patients with EGFR mutations-a review. *Translational lung cancer research*, 4: 67-81.
80. Lievre A, Bachet JB, Le Corre D, Boige V, Landi B, Emile JF, Cote JF, Tomasic G, Penna C, Ducreux M, Rougier P, Penault-Llorca F, Laurent-Puig P. (2006) KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. *Cancer research*, 66: 3992-3995.
81. Di Nicolantonio F, Martini M, Molinari F, Sartore-Bianchi A, Arena S, Saletti P, De Dosso S, Mazzucchelli L, Frattini M, Siena S, Bardelli A. (2008) Wild-type BRAF is required for response to panitumumab or cetuximab in metastatic colorectal cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 26: 5705-5712.
82. Berns K, Horlings HM, Hennessy BT, Madiredjo M, Hijmans EM, Beelen K, Linn SC, Gonzalez-Angulo AM, Stemke-Hale K, Hauptmann M, Beijersbergen RL, Mills GB, van de Vijver MJ, Bernards R. (2007) A functional genetic approach identifies the PI3K pathway as a major determinant of trastuzumab resistance in breast cancer. *Cancer cell*, 12: 395-402.
83. Ring A, Dowsett M. (2004) Mechanisms of tamoxifen resistance. *Endocrine-related cancer*, 11: 643-658.
84. Nowell PC. (1976) The clonal evolution of tumor cell populations. *Science*, 194: 23-28.
85. Greaves M, Maley CC. (2012) Clonal evolution in cancer. *Nature*, 481: 306-313.
86. Yates LR, Campbell PJ. (2012) Evolution of the cancer genome. *Nature reviews. Genetics*, 13: 795-806.
87. Gerlinger M, Horswell S, Larkin J, Rowan AJ, Salm MP, Varela I, Fisher R, McGranahan N, Matthews N, Santos CR, Martinez P, Phillimore B, Begum S, Rabinowitz A, Spencer-Dene B, Gulati S, Bates PA, Stamp G, Pickering L, Gore M, Nicol DL, Hazell S, Futreal PA, Stewart A, Swanton C. (2014) Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nature genetics*, 46: 225-233.
88. Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A, Tarpey P, Varela I, Phillimore B, Begum S, McDonald NQ, Butler A, Jones D, Raine K, Latimer C, Santos CR, Nohadani M, Eklund AC, Spencer-Dene B, Clark G, Pickering L, Stamp G, Gore M, Szallasi Z,

- Downward J, Futreal PA, Swanton C. (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *The New England journal of medicine*, 366: 883-892.
89. Bashashati A, Ha G, Tone A, Ding J, Prentice LM, Roth A, Rosner J, Shumansky K, Kalloger S, Senz J, Yang W, McConechy M, Melnyk N, Anglesio M, Luk MT, Tse K, Zeng T, Moore R, Zhao Y, Marra MA, Gilks B, Yip S, Huntsman DG, McAlpine JN, Shah SP. (2013) Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. *The Journal of pathology*, 231: 21-34.
 90. Zhang J, Fujimoto J, Wedge DC, Song X, Seth S, Chow CW, Cao Y, Gumbs C, Gold KA, Kalhor N, Little L, Mahadeshwar H, Moran C, Protopopov A, Sun H, Tang J, Wu X, Ye Y, William WN, Lee JJ, Heymach JV, Hong WK, Swisher S, Wistuba II, Futreal PA. (2014) Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science*, 346: 256-259.
 91. Yates LR, Gerstung M, Knappskog S, Desmedt C, Gundem G, Van Loo P, Aas T, Alexandrov LB, Larsimont D, Davies H, Li Y, Ju YS, Ramakrishna M, Haugland HK, Lilleng PK, Nik-Zainal S, McLaren S, Butler A, Martin S, Glodzik D, Menzies A, Raine K, Hinton J, Jones D, Mudie LJ, Jiang B, Vincent D, Greene-Colozzi A, Adnet PY, Fatima A, Maetens M, Ignatiadis M, Stratton MR, Sotiriou C, Richardson AL, Lonning PE, Wedge DC, Campbell PJ. (2015) Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nature medicine*, 21: 751-759.
 92. Lee JY, Yoon JK, Kim B, Kim S, Kim MA, Lim H, Bang D, Song YS. (2015) Tumor evolution and intratumor heterogeneity of an epithelial ovarian cancer investigated using next-generation sequencing. *BMC cancer*, 15: 85.
 93. Schwarz RF, Ng CK, Cooke SL, Newman S, Temple J, Piskorz AM, Gale D, Sayal K, Murtaza M, Baldwin PJ, Rosenfeld N, Earl HM, Sala E, Jimenez-Linan M, Parkinson CA, Markowitz F, Brenton JD. (2015) Spatial and temporal heterogeneity in high-grade serous ovarian cancer: a phylogenetic analysis. *PLoS medicine*, 12: e1001789.
 94. Sanger F, Nicklen S, Coulson AR. (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74: 5463-5467.

95. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, Hood LE. (1986) Fluorescence detection in automated DNA sequence analysis. *Nature*, 321: 674-679.
96. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IM, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DM, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Chiara ECM, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S, Ivanov DV, Johnson MQ, James T, Huw Jones TA, Kang GD, Kerelska TH, Kersey AD, Khrebtukova I, Kindwall AP, Kingsbury Z, Kokko-Gonzales PI, Kumar A, Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW, Newington T, Ning Z, Ling Ng B, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike AC, Chris Pinkard D, Pliskin DP, Podhasky J, Quijano VJ, Raczy C, Rae VH, Rawlings SR, Chiva Rodriguez A, Roe PM, Rogers J, Rogert Bacigalupo MC, Romanov N, Romieu A, Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA, Ernest Sohna Sohna J, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, Vandevondele S, Verhovsky Y, Virk SM, Wakelin S, Walcott GC, Wang J, Worsley GJ, Yan J, Yau L, Zuerlein M, Mullikin JC, Hurles ME, McCooke NJ, West JS, Oaks FL, Lundberg PL, Klenerman D, Durbin R, Smith AJ. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456: 53-59.

97. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fidanza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT, Bustillo J. (2011) An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, 475: 348-352.
98. Korlach J, Bjornson KP, Chaudhuri BP, Cicero RL, Flusberg BA, Gray JJ, Holden D, Saxena R, Wegener J, Turner SW. (2010) Real-time DNA sequencing from single polymerase molecules. *Methods in enzymology*, 472: 431-455.
99. Jain M, Fiddes IT, Miga KH, Olsen HE, Paten B, Akeson M. (2015) Improved data analysis for the MinION nanopore sequencer. *Nature methods*, 12: 351-356.
100. Ellingford JM, Barton S, Bhaskar S, Williams SG, Sergouniotis PI, O'Sullivan J, Lamb JA, Perveen R, Hall G, Newman WG, Bishop PN, Roberts SA, Leach R, Tearle R, Bayliss S, Ramsden SC, Nemeth AH, Black GC. (2016) Whole Genome Sequencing Increases Molecular Diagnostic Yield Compared with Current Diagnostic Testing for Inherited Retinal Disease. *Ophthalmology*, 123: 1143-1150.
101. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, Ercan-Sencicek AG, DiLullo NM, Parikshak NN, Stein JL, Walker MF, Ober GT, Teran NA, Song Y, El-Fishawy P, Murtha RC, Choi M, Overton JD, Bjornson RD, Carriero NJ, Meyer KA, Bilguvar K, Mane SM, Sestan N, Lifton RP, Gunel M, Roeder K, Geschwind DH, Devlin B, State MW. (2012) De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*, 485: 237-241.
102. Castera L, Krieger S, Rousselin A, Legros A, Baumann JJ, Bruet O, Brault B, Fouillet R, Goardon N, Letac O, Baert-Desurmont S, Tinat J, Bera O, Dugast C, Berthet P, Polycarpe F, Layet V, Hardouin A, Frebourg T, Vaur D. (2014) Next-generation sequencing for the diagnosis of hereditary breast and ovarian cancer using genomic capture targeting multiple candidate genes. *European journal of human genetics : EJHG*, 22: 1305-1313.
103. Best MG, Sol N, Kooi I, Tannous J, Westerman BA, Rustenburg F, Schellen P, Verschueren H, Post E, Koster J, Ylstra B, Ameziane N, Dorsman J, Smit EF, Verheul HM, Noske DP, Reijneveld JC, Nilsson RJA, Tannous BA, Wesseling P,

- Wurdinger T. (2015) RNA-Seq of Tumor-Educated Platelets Enables Blood-Based Pan-Cancer, Multiclass, and Molecular Pathway Cancer Diagnostics. *Cancer cell*, 28: 666-676.
104. Pritchard CC, Cheng HH, Tewari M. (2012) MicroRNA profiling: approaches and considerations. *Nature reviews. Genetics*, 13: 358-369.
 105. Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, Brown GD, Gojis O, Ellis IO, Green AR, Ali S, Chin SF, Palmieri C, Caldas C, Carroll JS. (2012) Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*, 481: 389-393.
 106. Delpu Y, Cordelier P, Cho WC, Torrisani J. (2013) DNA methylation and cancer diagnosis. *International journal of molecular sciences*, 14: 15029-15058.
 107. Pongor LS, Vera R, Ligeti B. (2014) Fast and sensitive alignment of microbial whole genome sequencing reads to large sequence datasets on a desktop PC: application to metagenomic datasets and pathogen identification. *PloS one*, 9: e103441.
 108. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, Zhang Y, Ye K, Jun G, Fritz MH, Konkel MK, Malhotra A, Stutz AM, Shi X, Casale FP, Chen J, Hormozdiari F, Dayama G, Chen K, Malig M, Chaisson MJP, Walter K, Meiers S, Kashin S, Garrison E, Auton A, Lam HYK, Mu XJ, Alkan C, Antaki D, Bae T, Cerveira E, Chines P, Chong Z, Clarke L, Dal E, Ding L, Emery S, Fan X, Gujral M, Kahveci F, Kidd JM, Kong Y, Lammeijer EW, McCarthy S, Flicek P, Gibbs RA, Marth G, Mason CE, Menelaou A, Muzny DM, Nelson BJ, Noor A, Parrish NF, Pendleton M, Quitadamo A, Raeder B, Schadt EE, Romanovitch M, Schlattl A, Sebra R, Shabalina AA, Untergasser A, Walker JA, Wang M, Yu F, Zhang C, Zhang J, Zheng-Bradley X, Zhou W, Zichner T, Sebat J, Batzer MA, McCarroll SA, Mills RE, Gerstein MB, Bashir A, Stegle O, Devine SE, Lee C, Eichler EE, Korb J. (2015) An integrated map of structural variation in 2,504 human genomes. *Nature*, 526: 75-81.
 109. Cock PJ, Fields CJ, Goto N, Heuer ML, Rice PM. (2010) The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic acids research*, 38: 1767-1771.
 110. Ewing B, Green P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome research*, 8: 186-194.

111. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA. (2013) From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current protocols in bioinformatics*, 43: 11 10 11-33.
112. Bolger AM, Lohse M, Usadel B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30: 2114-2120.
113. Smith TF, Waterman MS. (1981) Identification of common molecular subsequences. *Journal of molecular biology*, 147: 195-197.
114. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990) Basic local alignment search tool. *Journal of molecular biology*, 215: 403-410.
115. Li H, Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25: 1754-1760.
116. Li H, Durbin R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 26: 589-595.
117. Langmead B, Salzberg SL. (2012) Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9: 357-359.
118. Wang Z, Gerstein M, Snyder M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics*, 10: 57-63.
119. Trapnell C, Pachter L, Salzberg SL. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, 25: 1105-1111.
120. Kim D, Langmead B, Salzberg SL. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nature methods*, 12: 357-360.
121. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29: 15-21.
122. Ghosh S, Chan CK. (2016) Analysis of RNA-Seq Data Using TopHat and Cufflinks. *Methods in molecular biology*, 1374: 339-361.
123. Li B, Dewey CN. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics*, 12: 323.
124. Robinson MD, McCarthy DJ, Smyth GK. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26: 139-140.

125. Love MI, Huber W, Anders S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15: 550.
126. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25: 2078-2079.
127. Koboldt DC, Larson DE, Wilson RK. (2013) Using VarScan 2 for Germline Variant Calling and Somatic Mutation Detection. *Current protocols in bioinformatics*, 44: 15 14 11-17.
128. Larson DE, Abbott TE, Wilson RK. (2014) Using SomaticSniper to Detect Somatic Single Nucleotide Variants. *Current protocols in bioinformatics*, 45: 15 15 11-18.
129. Fan Y, Xi L, Hughes DS, Zhang J, Futreal PA, Wheeler DA, Wang W. (2016) MuSE: accounting for tumor heterogeneity using a sample-specific error model improves sensitivity and specificity in mutation calling from sequencing data. *Genome biology*, 17: 178.
130. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. (2012) Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics*, 28: 1811-1817.
131. Kroigard AB, Thomassen M, Laenkholtm AV, Kruse TA, Larsen MJ. (2016) Evaluation of Nine Somatic Variant Callers for Detection of Somatic Mutations in Exome and Targeted Deep Sequencing Data. *PloS one*, 11: e0151664.
132. Cai L, Yuan W, Zhang Z, He L, Chou KC. (2016) In-depth comparison of somatic point mutation callers based on different tumor next-generation sequencing depth data. *Scientific reports*, 6: 36540.
133. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome research*, 22: 568-576.
134. Favero F, Joshi T, Marquard AM, Birkbak NJ, Krzystanek M, Li Q, Szallasi Z, Eklund AC. (2015) Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Annals of oncology : official journal of the European Society for Medical Oncology*, 26: 64-70.
135. Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, Laird PW, Onofrio RC, Winckler W, Weir BA, Beroukhim R, Pellman D, Levine DA, Lander

- ES, Meyerson M, Getz G. (2012) Absolute quantification of somatic DNA alterations in human cancer. *Nature biotechnology*, 30: 413-421.
136. Saad ED, Katz A. (2009) Progression-free survival and time to progression as primary end points in advanced breast cancer: often used, sometimes loosely defined. *Annals of oncology : official journal of the European Society for Medical Oncology*, 20: 460-464.
 137. Dinse GE, Lagakos SW. (1982) Nonparametric estimation of lifetime and disease onset distributions from incomplete observations. *Biometrics*, 38: 921-932.
 138. Mantel N. (1966) Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer chemotherapy reports*, 50: 163-170.
 139. Gyorffy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q, Szallasi Z. (2010) An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast cancer research and treatment*, 123: 725-731.
 140. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, 6: 80-92.
 141. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic acids research*, 29: 308-311.
 142. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, Jang W, Katz K, Ovetsky M, Riley G, Sethi A, Tully R, Villamarin-Salomon R, Rubinstein W, Maglott DR. (2016) ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic acids research*, 44: D862-868.
 143. Francis Blokzijl RJ, Ruben Van Boxtel, Edwin Cuppen. (2016) MutationalPatterns: an integrative R package for studying patterns in base substitution catalogues. *bioRxiv*.
 144. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G. (2013) Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature biotechnology*, 31: 213-219.

145. Gautier L, Cope L, Bolstad BM, Irizarry RA. (2004) affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*, 20: 307-315.
146. Sing T, Sander O, Beerenwinkel N, Lengauer T. (2005) ROCR: visualizing classifier performance in R. *Bioinformatics*, 21: 3940-3941.
147. Pongor L, Kormos M, Hatzis C, Pusztai L, Szabo A, Gyorffy B. (2015) A genome-wide approach to link genotype to clinical outcome by utilizing next generation sequencing and gene chip data of 6,697 breast cancer patients. *Genome medicine*, 7: 104.
148. Thorvaldsdottir H, Robinson JT, Mesirov JP. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*, 14: 178-192.
149. Jiang T, Shi W, Wali VB, Pongor LS, Li C, Lau R, Gyorffy B, Lifton RP, Symmans WF, Pusztai L, Hatzis C. (2016) Predictors of Chemosensitivity in Triple Negative Breast Cancer: An Integrated Genomic Analysis. *PLoS medicine*, 13: e1002193.
150. Pereira B, Chin SF, Rueda OM, Vollan HK, Provenzano E, Bardwell HA, Pugh M, Jones L, Russell R, Sammut SJ, Tsui DW, Liu B, Dawson SJ, Abraham J, Northen H, Peden JF, Mukherjee A, Turashvili G, Green AR, McKinney S, Oloumi A, Shah S, Rosenfeld N, Murphy L, Bentley DR, Ellis IO, Purushotham A, Pinder SE, Borresen-Dale AL, Earl HM, Pharoah PD, Ross MT, Aparicio S, Caldas C. (2016) The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nature communications*, 7: 11479.
151. Takaku M, Grimm SA, Wade PA. (2015) GATA3 in Breast Cancer: Tumor Suppressor or Oncogene? *Gene expression*, 16: 163-168.
152. Huang da W, Sherman BT, Lempicki RA. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*, 4: 44-57.
153. Nagy A, Pongor LS, Szabo A, Santarpia M, Gyorffy B. (2017) KRAS driven expression signature has prognostic power superior to mutation status in non-small cell lung cancer. *International journal of cancer*, 140: 930-937.
154. Harami-Papp H, Pongor LS, Munkacsy G, Horvath G, Nagy AM, Ambrus A, Hauser P, Szabo A, Tretter L, Gyorffy B. (2016) TP53 mutation hits energy metabolism and increases glycolysis in breast cancer. *Oncotarget*, 7: 67183-67195.

155. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic acids research*, 44: D457-462.
156. Shubbar E, Kovacs A, Hajizadeh S, Parris TZ, Nemes S, Gunnarsdottir K, Einbeigi Z, Karlsson P, Helou K. (2013) Elevated cyclin B2 expression in invasive breast carcinoma is associated with unfavorable clinical outcome. *BMC cancer*, 13: 1.
157. Karra H, Repo H, Ahonen I, Loyttyneimi E, Pitkanen R, Lintunen M, Kuopio T, Soderstrom M, Kronqvist P. (2014) Cdc20 and securin overexpression predict short-term breast cancer survival. *British journal of cancer*, 110: 2905-2913.
158. Daniel J, Coulter J, Woo JH, Wilsbach K, Gabrielson E. (2011) High levels of the Mps1 checkpoint protein are protective of aneuploidy in breast cancer cells. *Proceedings of the National Academy of Sciences of the United States of America*, 108: 5384-5389.
159. Percy MJ, Myrie KA, Neeley CK, Azim JN, Ethier SP, Petty EM. (2000) Expression and mutational analyses of the human MAD2L1 gene in breast cancer cells. *Genes, chromosomes & cancer*, 29: 356-362.
160. Michalides R, van Tinteren H, Balkenende A, Vermorken JB, Benraadt J, Huldij J, van Diest P. (2002) Cyclin A is a prognostic indicator in early stage breast cancer with and without tamoxifen treatment. *British journal of cancer*, 86: 402-408.
161. Myrie KA, Percy MJ, Azim JN, Neeley CK, Petty EM. (2000) Mutation and expression analysis of human BUB1 and BUB1B in aneuploid breast cancer cell lines. *Cancer letters*, 152: 193-199.
162. Bonin S, Brunetti D, Benedetti E, Gorji N, Stanta G. (2006) Expression of cyclin-dependent kinases and CDC25a phosphatase is related with recurrences and survival in women with peri- and post-menopausal breast cancer. *Virchows Archiv : an international journal of pathology*, 448: 539-544.
163. Sonoda H, Inoue H, Ogawa K, Utsunomiya T, Masuda TA, Mori M. (2006) Significance of skp2 expression in primary breast cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 12: 1215-1220.
164. Bonte D, Lindvall C, Liu H, Dykema K, Furge K, Weinreich M. (2008) Cdc7-Dbf4 kinase overexpression in multiple cancers and tumor cell lines is correlated with p53 inactivation. *Neoplasia*, 10: 920-931.

165. Gonzalez MA, Pinder SE, Callagy G, Vowler SL, Morris LS, Bird K, Bell JA, Laskey RA, Coleman N. (2003) Minichromosome maintenance protein 2 is a strong independent prognostic marker in breast cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 21: 4306-4313.
166. Vaziri C, Saxena S, Jeon Y, Lee C, Murata K, Machida Y, Wagle N, Hwang DS, Dutta A. (2003) A p53-dependent checkpoint pathway prevents rereplication. *Molecular cell*, 11: 997-1008.
167. Keyomarsi K, Tucker SL, Buchholz TA, Callister M, Ding Y, Hortobagyi GN, Bedrosian I, Knickerbocker C, Toyofuku W, Lowe M, Herliczek TW, Bacus SS. (2002) Cyclin E and survival in patients with breast cancer. *The New England journal of medicine*, 347: 1566-1575.
168. Weichert W, Kristiansen G, Winzer KJ, Schmidt M, Gekeler V, Noske A, Muller BM, Niesporek S, Dietel M, Denkert C. (2005) Polo-like kinase isoforms in breast cancer: expression patterns and prognostic implications. *Virchows Archiv : an international journal of pathology*, 446: 442-450.
169. Verlinden L, Vanden Bempt I, Eelen G, Drijkoningen M, Verlinden I, Marchal K, De Wolf-Peeters C, Christiaens MR, Michiels L, Bouillon R, Verstuyf A. (2007) The E2F-regulated gene Chk1 is highly expressed in triple-negative estrogen receptor /progesterone receptor /HER-2 breast carcinomas. *Cancer research*, 67: 6574-6581.
170. Rasche L, Chavan SS, Stephens OW, Patel PH, Tytarenko R, Ashby C, Bauer M, Stein C, Deshpande S, Wardell C, Buzder T, Molnar G, Zangari M, van Rhee F, Thanendrarajan S, Schinke C, Epstein J, Davies FE, Walker BA, Meissner T, Barlogie B, Morgan GJ, Weinhold N. (2017) Spatial genomic heterogeneity in multiple myeloma revealed by multi-region sequencing. *Nature communications*, 8: 268.
171. Li S, Garrett-Bakelman FE, Chung SS, Sanders MA, Hricik T, Rapaport F, Patel J, Dillon R, Vijay P, Brown AL, Perl AE, Cannon J, Bullinger L, Luger S, Becker M, Lewis ID, To LB, Delwel R, Lowenberg B, Dohner H, Dohner K, Guzman ML, Hassane DC, Roboz GJ, Grimwade D, Valk PJ, D'Andrea RJ, Carroll M, Park CY, Neuberg D, Levine R, Melnick AM, Mason CE. (2016) Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia. *Nature medicine*, 22: 792-799.

172. Southey MC, Winship I, Nguyen-Dumont T. (2016) PALB2: research reaching to clinical outcomes for women with breast cancer. *Hereditary cancer in clinical practice*, 14: 9.
173. Antoniou AC, Casadei S, Heikkinen T, Barrowdale D, Pylkas K, Roberts J, Lee A, Subramanian D, De Leeneer K, Fostira F, Tomiak E, Neuhausen SL, Teo ZL, Khan S, Aittomaki K, Moilanen JS, Turnbull C, Seal S, Mannermaa A, Kallioniemi A, Lindeman GJ, Buys SS, Andrulis IL, Radice P, Tondini C, Manoukian S, Toland AE, Miron P, Weitzel JN, Domchek SM, Poppe B, Claes KB, Yannoukakos D, Concannon P, Bernstein JL, James PA, Easton DF, Goldgar DE, Hopper JL, Rahman N, Peterlongo P, Nevanlinna H, King MC, Couch FJ, Southey MC, Winqvist R, Foulkes WD, Tischkowitz M. (2014) Breast-cancer risk in families with mutations in PALB2. *The New England journal of medicine*, 371: 1651-1652.

11. Saját publikációk jegyzéke

11.1. Disszertációhoz kapcsolódó publikációk jegyzéke

1. **Pongor LS**, Kormos M, Hatzis C, Pusztai L, Szabo A, Gyorffy B. (2015) A genome-wide approach to link genotype to clinical outcome by utilizing next generation sequencing and gene chip data of 6,697 breast cancer patients. *Genome Medicine*, 7: 104. **IF=5,846**
2. Jiang T, Shi W, Wali VB, **Pongor LS**, Li C, Lau R, Gyorffy B, Lifton RP, Symmans WF, Pusztai L, Hatzis C. (2016) Predictors of Chemosensitivity in Triple Negative Breast Cancer: An Integrated Genomic Analysis. *PLoS Medicine*, 13: e1002193. **IF= 11,862**
3. Harami-Papp H*, **Pongor LS***, Munkacsy G, Horvath G, Nagy AM, Ambrus A, Hauser P, Szabo A, Tretter L, Gyorffy B. (2016) TP53 mutation hits energy metabolism and increases glycolysis in breast cancer. *Oncotarget*, 7: 67183-67195. **IF= 5,168**
4. Nagy A, **Pongor LS**, Szabo A, Santarpia M, Gyorffy B. (2017) KRAS driven expression signature has prognostic power superior to mutation status in non-small cell lung cancer. *International Journal of Cancer*, 140: 930-937. **IF=6,513**
5. **Pongor LS**, Harami-Papp H, Mehes E, Czirok A, Gyorffy B. (2017) Cell Dispersal Influences Tumor Heterogeneity and Introduces a Bias in NGS Data Interpretation. *Scientific Reports*, 7: 7358. **IF= 4,259**

11.2. Disszertációtól független publikációk jegyzéke

1. **Pongor LS**, Pinter F, Petak I. (2013) HeurAA: accurate and fast detection of genetic variations with a novel heuristic amplicon aligner program for next generation sequencing. *PloS One*, 8: e54294. **IF= 3.534**

2. **Pongor LS**, Vera R, Ligeti B. (2014) Fast and sensitive alignment of microbial whole genome sequencing reads to large sequence datasets on a desktop PC: application to metagenomic datasets and pathogen identification. PloS One, 9: e103441. **IF= 3.234**

3. Cai W, Xiong Chen Z, Rane G, Satendra Singh S, Choo Z, Wang C, Yuan Y, Zea Tan T, Arfuso F, Yap CT, **Pongor LS**, Yang H, Lee MB, Cher Goh B, Sethi G, Benoukraf T, Tergaonkar V, Prem Kumar A. (2017) Wanted DEAD/H or Alive: Helicases Winding Up in Cancers. Journal of the National Cancer Institute, 109. **IF= 12.589**

*Megosztott első szerző

12. Köszönetnyilvánítás

Elsőként témavezetőmnek, Dr. Györffy Baláznak szeretnék köszönetet mondani, aki optimális háttérrel biztosított munkám elvégzéséhez, és javaslataival és tanácsaival végig segítette munkámat.

Szeretnék köszönetet mondani az II. Sz. Gyermekgyógyászati Klinika igazgatójának, Prof. Dr. Szabó Andrásnak, hogy kutatásaimat az II. Sz. Gyermekgyógyászati Klinikán végezhettem.

Köszönettel tartozom a munkacsoport tagjainak, Dr. Sztupinszki Zsófia és Nagy Ádám PhD hallgató társaimnak, Dr. Harami-Papp Hajnalka és Dr. Munkácsy Gyöngyi laboratóriumban dolgozó munkatársaimnak, illetve a csoport többi tagjának, akik munkámat segítették.

Továbbá köszönettel tartozom a tudományos együttműködőknek – Dr. Pusztai Lajosnak és Dr. Hatzis Christosnak (Department of Medicine, Yale School of Medicine, Yale University), Méhes Elődnek (Biológiai Fizika Tanszék, Eötvös Loránd Tudományegyetem), valamint Dr. Pete Imrének és Dr. Vereczkey Ildikónak (Országos Onkológiai Intézet), - akik munkája nélkülözhetetlen volt kutatásom során.

Köszönettel tartozom a szüleimnek és családomnak, akik támogattak a munkám során.

Végül köszönetet mondok feleségemnek, aki jó tanácsokkal, sok segítséggel és türelemmel támogatta munkámat.